JUL 18 1949

# The Journal of Experimental Education

A periodical report of scientific investigations relating to child development, curriculum, learning, teaching, supervision, measurements, statistics, and experimental techniques.

Volume XVII | March, 1949 | Number 3

## CONTENTS

PAGE

Projections of Three Types of Factor Pattern: *Chester W. Harris* ..... 335

The Development of a Technique for the Measurement of Social-Emotional Climate in Classrooms: *John Withall* ..... 347

An Experimental Study of the Effectiveness of Filmstrips in Teaching Geometry: *Donovan A. Johnson* ..... 363

An Experimental Evaluation of the Relative Effectiveness of Certain Audio-Visual Aids in Vocational Agriculture: *Ralph R. Bentley* ..... 373

Note on Establishing Norms in a Large City System: *J. A. Keats* ..... 383

A Graphical Representation of Test Reliability: *John A. Jarvis* ..... 387

$5.00 a Year | Published Quarterly | $1.50 a Copy

Published by Dunbar Publications, Inc., 114 S. Carroll Street, Madison 3, Wisconsin.

Entered as second-class matter October 17, 1938 at the post office at Madison, Wisconsin under the act of March 3, 1879.

## EDITORIAL BOARD

**A. S. Barr, Chairman, Professor of Education, University of Wisconsin, Madison 6, Wis.**

Arthur T. Jersild, Professor of Education, Teachers College, Columbia University, New York City. Editorially responsible for materials on child welfare, guidance, and development, published each December.

Palmer O. Johnson, Professor of Education, University of Minnesota, Minneapolis, Minnesota. Editorially responsible for materials on measurements, statistics, and methods of experimental research, published each March.

H. H. Remmers, Professor of Educational Psychology, Director Division Educational Reference, Purdue University, Lafayette, Indiana. Editorially responsible for materials on learning, teaching and supervision, published each September.

J. Wayne Wrightstone, Assistant Director, Bureau of Reference, Research and Statistics, Board of Education of the City of New York, 110 Livingston Street, Brooklyn, New York. Editorially responsible for materials on curriculum construction, published each June.

## CONTRIBUTING EDITORS

Emmett A. Betts, Research Professor and Director of Reading Clinic, Pennsylvania State College, State College, Pennsylvania.

Gilbert L. Betts, Supervisor of Graduate Research in Education, Colorado State College, Fort Collins, Colorado.

Leo J. Brueckner, Professor of Education, University of Minnesota, Minneapolis, Minnesota.

Oscar K. Buros, Associate Professor of Education, Rutgers University, New Brunswick, New Jersey.

Guy T. Buswell, Professor of Educational Psychology, University of Chicago, Chicago, Illinois.

Harold D. Carter, Associate Professor of Education, University of California, Berkeley 4, California.

Leslie L. Chisholm, Associate Professor of Education, State College of Washington, Pullman, Washington.

Herbert S. Conrad, Technical Consultant, College Entrance Examination Board, Princeton, New Jersey.

Stephen M. Corey, Professor of Educational Psychology, University of Chicago, Chicago, Illinois.

Edward E. Cureton, Secretary–Treasurer, Richardson, Bellows, Henry & Company, Inc., New York 4, New York.

Robert A. Davis, Professor of Education, Director of Bureau of Educational Research, University of Colorado, Boulder, Colorado.

Hiel R. Douglass, Director of College of Education, University of Colorado, Boulder, Colorado.

Harold A. Edgerton, Director, Occupational Opportunities Service, Professor of Psychology, Ohio State University, Columbus 10, Ohio.

John C. Flanagan, Professor of Psychology, University of Pittsburgh, Pittsburgh, Pennsylvania.

Carter V. Good, Dean, Teachers College, University of Cincinnati, Cincinnati 21, Ohio.

Robert W. B. Jackson, Assistant Professor of Educational Research; Assistant Director, Department of Educational Research, Ontario College of Education, University of Toronto, Toronto, Canada.

Harold E. Jones, Professor of Psychology and Director, Institute of Child Welfare, University of California, Berkeley 4, California.

Neal Kaye, Professor of Education and Lecturer in Human Relations, University of California, Berkeley, California.

D. Welty Lefever, Professor of Education, University of Southern California, Los Angeles, California.

Edward A. Lincoln, Consulting Psychologist, Halifax, Massachusetts.

Irving Lorge, Professor of Education, Executive Officer, Institute of Psychological Research, Teachers College, Columbia University, New York 27, New York.

A. R. Mead, Director of Educational Research, University of Florida, 330 P. K. Yonge Building, Gainesville, Florida.

T. E. Newland, Lt. Comdr., USNR, 2703 Wisconsin Avenue, N. W., Washington 7, D. C.

C. W. Odell, Professor of Education, University of Illinois, Urbana, Illinois.

Willard C. Olson, Professor of Education, Director of Research in Child Development, University of Michigan, Ann Arbor, Michigan.

W. E. Peik, Dean and Professor of Education, University of Minnesota, Minneapolis, Minnesota.

Walworth R. Plumb, Director Student Personnel Service, University of Minnesota (Branch), Duluth, Minnesota.

S. L. Pressey, Professor of Educational Psychology, Ohio State University, Columbus, Ohio.

Clarence R. Ragsdale, Professor of Education, University of Wisconsin, Madison, Wisconsin.

William Reitz, Associate Professor of Education, College of Education Examiner, Wayne University, Detroit 2, Michigan.

Henry D. Rinsland, Professor of Education and Director of Educational Research, The University of Oklahoma, Norman, Oklahoma.

Robert T. Rock, Jr., Professor of Psychology, Head of Dept. of Psychology, Graduate School, Fordham University, New York City.

Philip J. Rulon, Professor of Education, Harvard Graduate School of Education, Cambridge 38, Mass.

Douglas E. Scates, Professor of Education, Duke University, Durham, North Carolina.

Harold Seashore, Director, Test Division, The Psychological Corporation, New York 18, New York.

David Segel, Educational Consultant, Specialist in Tests and Measurements, Federal Security Agency, U. S. Office of Education, Washington, D. C.

Paul W. Terry, Professor of Educational Psychology, University of Alabama, University, Alabama.

Helen Thompson, Associate Attending Psychologist, New York Post–Graduate Hospital, 303 East 20th Street, New York 3, New York.

Robert L. Thorndike, Associate Professor of Education, Teachers College, Columbia University, New York City.

Herbert A. Toops, Professor of Psychology, Ohio State University, Columbus, Ohio.

Maurice E. Troyer, Director, Bureau of School Service, Syracuse University, Syracuse 10, New York.

Helen M. Walker, Professor of Education, Teachers College, Columbia University, New York City.

Beth L. Wellman, Professor of Psychology, Child Welfare Research Station, State University of Iowa, Iowa City, Iowa.

Guy M. Wilson, Professor of Education, Emeritus, Boston University, Boston, Massachusetts.

Paul A. Witty, Professor of Education, Director of Psycho-Educational Clinic, School of Education, Northwestern University, Evanston, Illinois.

Ernest R. Wood, Professor of Education, New York University, New York City.

D. A. Worcester, Chairman Department of Educational Psychology and Measurement, University of Nebraska, Lincoln, Nebraska.



LITHO PRODUCTIONS

MADISON WIS

Journal of Experimental Education

Volume XVII March, 1949 Number 3

# PROJECTIONS OF THREE TYPES OF FACTOR PATTERN

CHESTER W. HARRIS *

University of Wisconsin

In transforming any given correlation matrix, R, into a primary factor solution, two steps are commonly taken. The first is to transform the n by n correlation matrix, of rank m, into an arbitrary orthogonal factor matrix of size n by m, such as the centroid matrix, F. The second is to rotate this factor matrix into the desired primary factor pattern. This second step obviously depends upon a concept of the desired primary factor pattern. An indefinite number of rotations might be made, and consequently one has no notion of when to cease this operation unless an exact step-by-step specification of the number and character of the rotations to be made is set up beforehand, or a criterion for recognizing the primary solution is available. An attempt is made in this paper to give a geometric formulation of the primary factor solution that provides such a criterion for certain types of desired factor pattern. One approach to the translation of this into an analytic criterion will then be suggested.

The method to be used is simple in outline. A particular type of figure plus its analogues in higher-dimensional space will be postulated as the model for the primary solution. This figure is the tetrahedron, a three-dimensional figure formed by four triangles, which consequently has four apexes and six edges. The projections of this figure and of its analogues into a space of one less dimension will then be specified. Next, certain factor patterns will be specified as desired ones, and the projections of the centroids they generate will be shown to be consistent with the projections of certain regions of the postulated model. Since certain of these projected regions are represented as points and lines, one avenue to the development of an analytic criterion may be opened up.

## The Model and its Projections

An essential characteristic of the extended-vectors representation of a set of centroids is that it projects the test vectors into an (m - 1)-space.[1] In order to relate the postulated model to the extended-vectors representation of certain types of factor pattern, it will therefore be necessary to develop both the m-dimensional model and its projection into (m - 1)-space. The tetrahedron has been described above. Specify one apex as the fixed origin to correspond with the factor problem in which the test vectors have a common origin. Let the three planes that meet at this fixed origin each be defined by a large number of test vectors, but let the "interior" of the tetrahedron be hollow. If each test vector is now extended into the plane opposite the origin, these test vectors project as points that define three lines arranged in the form of a triangle. This is, of course, the familiar extended-vectors representation of an "ideal" three-factor primary pattern.[2]

Two principles are to be established. For the 3-space model (the tetrahedron) each of the three coordinate planes projects as a line, which is defined by the points representing test vectors that lie in that plane. The three lines have a special arrangement, since they form a triangle. In general, then, the projection of the coordinate hyperplanes of an m-space figure will be required to yield the

* I am deeply indebted to Professor J. R. Mayor, Department of Mathematics, University of Wisconsin, for reading this paper in manuscript, and suggesting a number of changes. Any errors that remain are solely mine.

1. See L. L. Thurstone, Multiple-Factor Analysis (Chicago: University of Chicago Press, 1947), Chapter XI, for a description of this representation.

2. See Thurstone, op. cit., p. 228, for an illustration.

(m - 1)-space figure that constitutes a hyperplane of the m-space figure. By beginning with the triangle as a hyperplane of the tetrahedron, it is possible to define an indefinite series of projections of configurations that lie in higher-dimensional space. The triangle and the tetrahedron also give the principle for specifying the number of apexes, bounding hyperplanes, and intersections of bounding hyperplanes for any such configuration. If m equals the space within which the figure lies, then m + 1 gives the number of apexes, m + 1 also gives the number of bounding hyperplanes, and $\frac{(m+1)(m)}{2}$ gives the number of intersections of these bounding hyperplanes, pair by pair.

In order to define the 4-space hypertetrahedron, make use of the two principles given above. First, its projection in 3-space will be a tetrahedron. Because we wish to use the extended-vectors representation in dealing with the factor problem, however, this projection in 3-space will be regarded as a skeleton or outline of a tetrahedron. The distinctive characteristics of this skeleton will be given by six lines so arranged as to meet in four points in 3-space; these lines are, of course, defined by points that are the projections of the test vectors. In contrast to its projection, the 4-space figure should have five apexes, five bounding hyperplanes, and ten intersections of these. Again assume that coordinate test vectors define the four coordinate hyperplanes, i.e., those that meet at the special apex called the origin. The requirement is that each of these hyperplanes project as a hyperplane of the next lowest dimensional figure, i.e., as a three-sided figure lying in a plane, or a triangle. The skeleton of the tetrahedron, which is the projection of the 4-space hypertetrahedron, has four such bounding planes; each one therefore represents the projection of a hyperplane of the 4-space figure. But such a hyperplane is by definition a tetrahedron in which only three planes are coordinate with the origin and defined by test vectors; therefore each such hyperplane should project as a triangle. Defining the 4-space hypertetrahedron as five tetrahedrons so arranged as to meet in five points in 4-space is therefore consistent. Of these five tetrahedrons, only four are coordinate with the fixed origin and consequently are defined by test vectors. The fifth tetrahedron therefore is taken as the 3-space model of the projection of the test vectors defining this 4-space figure, just as the triangle opposite the origin is taken as the 2-space model of the projection of the test vectors defining a tetrahedron. Such a definition of the model and its projection may be generalized to a space of any dimensions.[3]

It is necessary to account for the six lines postulated in the projection of the 4-space hypertetrahedron in terms of planes within the 4-space figure. In general,

$$m + i = h + k$$

The intersection, i, of a (m - 1)-space hyperplane, h, and a plane k, in an m-space gives a line (1-space), thus:

$$m + i = (m - 1) + 2$$

and

$$i = 1$$

Similarly, any line, such as a test vector, intersects with such a hyperplane as a point. The test vectors therefore project into the (m - 1)-space as points. The test vectors defining coordinate hyperplanes project as points that define lines. These lines, therefore, represent planes of the coordinate hyperplanes. This is obvious in the projection of the three-factor problem for which the tetrahedron is the model, since for it the coordinate hyperplanes are simply the three planes that project as lines. The six planes, each represented by a line in the projection, of the 4-space figure present a problem, however. Each of the four coordinate hyperplanes of the 4-space hypertetrahedron is conceived as a tetrahedron. Because of the restriction of a common origin for the test vectors, however, test vectors may lie in only three of the four bounding planes of any of these tetrahedrons. The expectation might therefore be that three times four, or twelve, such planes would project as lines. An additional restriction is necessary. It is that the four coordinate tetrahedrons of the 4-space figure must be arranged in 4-space in a manner analogous to the arrangement of the four planes of a tetrahedron in 3-space. Each of these four coordinate tetrahedrons of the 4-space figure will therefore have a plane in common with every other such tetrahedron. This means that only one-half of these twelve planes, or six of them, are distinct in the arrangement of the four hyperplanes.

It can be demonstrated that each apex of the 4-space figure is required by these restrictions to be a point. This distinctive arrangement specifies that an apex is the common intersection of m of the m + 1 hyperplanes. The intersection of two hyperplanes, in m-space, gives a lower dimensional m - 2 hyperplane; of three hyperplanes, a lower dimensional m - 3 hyperplane; and so on. Therefore the intersection

3. Sommerville calls this model a simplex of m dimensions. Its projection would be, in his terminology, a simplex of m - 1 dimensions. See D. M. Y. Sommerville, An Introduction to the Geometry of N Dimensions (London: Methuen and Co., Ltd., 1929), Chapters I and VII.

of m hyperplanes in an m-space gives a figure of m - m dimensions, or a point. Since each apex of the m-space is a point, each pair of apexes may be joined by a straight line. The requirement of coordinate test vectors, however, restricts our interest to the m such lines that may join the special apex called the origin to each of the other m apexes. If these lines project into the (m - 1)-space, they will project as points. Now consider the projection of the m-space figure. The apexes of the projection will number exactly m; each such apex is a point and therefore represents the projection of a line of the m-space figure. Further, each such apex of the projection is the common intersection of m - 1 of the hyperplanes of that projection. Since each hyperplane of the projection represents a hyperplane of the m-space figure, it follows that each of these m lines represents the intersection in m-space of m - 1 hyperplanes of the m-space hypertetrahedron. The general theorem indicates that the intersection of m - 1 hyperplanes in an m-space defines a line. The apexes of the projection therefore represent certain special lines formed by the intersection of m - 1 of the coordinate hyperplanes of the m-space figure. The origin is a point on each of these m lines.

With the definition of the 4-space figure and its projection, the definition of higher-space figures becomes a matter of translating each one into the projection of the next higher one. A summary may be given as follows:

The Hypertetrahedron

| Space in which the hypertetrahedron lies | Number of apexes and of hyperplanes | Number of coordinate hyperplanes | Space in which each coordinate hyperplane lies |
|---|---|---|---|
| 3 | 4 | 3 | 2 |
| 4 | 5 | 4 | 3 |
| 5 | 6 | 5 | 4 |
| . | . | . | . |
| m | m+1 | m | m-1 |

The Projection

| Space in which the hypertetrahedron lies | Space in which the projection lies | Number of coordinate hyperplanes that project | Configuration of the projection of each hyperplane | Configuration of the total projection |
|---|---|---|---|---|
| 3 | 2 | 3 | one line in 1-space | a triangle |
| 4 | 3 | 4 | three lines in 2-space | a tetrahedron |
| 5 | 4 | 5 | six lines in 3-space | a 4-space hypertetrahedron |
| . | . | . | . . . | . . . |
| m | m-1 | m | $\frac{(m-1)(m-2)}{2}$ lines in (m-2)-space | a (m-1)-space hypertetrahedron |

## Representation on a Plane

In addition to defining the projections of the m-space configuration, it is necessary to find a way of representing each one on a plane. This is no problem for the projection of the three-factor problem, since its model, the tetrahedron, projects as a plane figure. The 4-space hypertetrahedron, however, projects as a 3-space figure--the skeleton or outline of a tetrahedron. This could be constructed in space, but the projection of any higher-space figure could not, since the projection of any one of these occupies at least a 4-space. A plane representation is also needed in dealing with the factor problem. To secure one, we may cast the shadow, so to speak, of the projection onto a plane surface. The diagrams of Figure 1 illustrate plane representations of the projection of a 4-space hypertetrahedron. Since m is four, four points are specified and then connected pair by pair with straight lines. Each of these diagrams is a familiar figure, since it is a line drawing of a tetrahedron in perspective. Although the six lines actually lie in a plane, they can be "seen" to outline a 3-space figure.

The reason for the three diagrams in Figure 1 can be explained readily. If the projection lies in a three or higher space, there is a large number of possible representations on a plane of the same figure. Although each such representations would be different from every other, in an absolute sense, there would be many that would show the same relative arrangement of lines and points. The three distinctive arrangements of the representation of the tetrahedron are shown as Figures 1a, 1b, and 1c. Three orthogonal axes may be placed in a 3-space. Each representation in Figure 1 gives the figure in perspective from a point on one of these three axes. The number of mutually orthogonal planes that can be erected in the space occupied by the projection gives the number of distinctive relative arrangements of the points and lines of the representation. Therefore, the projection of an m-space hypertetrahedron may be represented on a plane in $\frac{(m-1)(m-2)}{2}$ distinctive ways. Each such representation will also be seen to be a rotation through one dimension of of any other representation. Letters are used to designate the four apexes in the diagrams of Figure 1 so that this rotation may be noted.

By joining points with straight lines, therefore, the projection may be represented on a plane in a certain number of distinctive ways. The number of points is given by m, since this is the number of lines joining the origin and each of the other apexes of the m-space figure that project into the (m - 1)-space. The number of lines in the projection is given by $\frac{(m)(m-1)}{2}$, since this is the number of distinctive planes making up the m coordinate hyperplanes of the m-space hypertetrahedron that project into the (m - 1)-space. This is true because a plane in the m-space hypertetrahedron will be determined by the intersection of m - 2 hyperplanes. Since planes formed by the intersections of only the m coordinate hyperplanes will project, the number of such planes is given by the combination of m things taken m - 2 at a time, which reduces to $\frac{(m)(m-1)}{2}$. The projection of a 5-space hypertetrahedron thus may be represented on a plane by five points joined by ten straight lines; of a 6-space hypertetrahedron by six points joined by fifteen straight lines; and so on.

## Factor Patterns

A factor pattern will be defined as the coefficients of a set of linear equations describing the composition of the variables in terms of the factors. This is Holzinger's definition.[4] These coefficients also are the Cartesian coordinates of the variables with respect to the axes defined by the factors. A factor pattern may be represented as a matrix of size n by m, where n designates the number of variables and m the number of common factors. In this representation, uniqueness is disregarded. The term "primary factor pattern" will designate that the variables are described in terms of primary factors. Since the major problem of this paper is the geometric formulation of the primary factors, no attempt is made to define them a priori. However, the primary factors will be permitted to be oblique, as a general case, with orthogonal primaries as a special case. Three types of primary factor pattern will be considered in this paper. Each is defined in terms of the maximum complexity of any variable. Complexity refers to the number of non-zero coefficients in the linear equation describing the variable in terms of the factors. Thus a variable of complexity one will have only one such coefficient, a variable of complexity two, two such coefficients, etc. The cluster pattern will be defined as a pattern in which the maximum complexity of any variable is one, a 2-multiple-factor pattern as one in which the maximum complexity of any variable is two, and a 3-multiple-factor pattern as one in which the maximum complexity of any variable is three.[5] Fictitious examples of these three types of pattern may readily be constructed by specifying numerical values for the non-zero coefficients. The centroids they generate may then be com-

4. Karl J. Holzinger and Harry H. Harman, Factor Analysis (Chicago: University of Chicago Press, 1941), pp. 16-18.

puted and the vectors extended to unit correlation with the first centroid. Plots of the various pairs of the columns, other than the first, of this extended-vector matrix then give the complete extended-vectors representation of the pattern. Since the extended-vector representation of three-factor patterns is well known, fictitious patterns for a four-factor problem will be illustrated.[6]

The cluster pattern, when m equals four, may be represented:

| | $P_1$ | $P_2$ | $P_3$ | $P_4$ |
|---|---|---|---|---|
| A | + | 0 | 0 | 0 |
| B | 0 | + | 0 | 0 |
| C | 0 | 0 | + | 0 |
| D | 0 | 0 | 0 | + |

where $P_1$, $P_2$, $P_3$, $P_4$ designate the four primary axes, and A, B, C, D each designate a colinear cluster of tests. Each test in a given cluster is of complexity one, designated by the single plus sign in that row. The entries in the matrix of intercorrelations of the four primary axes, $\Phi$, need meet only the restrictions that are imposed upon a matrix of intercorrelations of unit vectors that are linearly independent; this permits the case of orthogonal primaries, or Holzinger's uni-factor pattern. The three extended-vectors diagrams of the centroids generated by such a factor pattern are given in Figure 2.[7]

It is not surprising that each diagram reveals only the extensions of the four clusters. Two points are of importance here. The four clusters have different positions relative to each other in each of the three diagrams. Figure 2b, for example, represents a rotation through one dimension of the (hypothetical) line joining clusters A and B as represented in Figure 2a. Figure 2c is related to Figure 2b and to Figure 2a in a similar fashion. These, then, represent the three distinctive views possible in the (m - 1)-space of the extended-vectors representation of a four-factor problem. Second, each of these diagrams suggests the projection of the assumed hypertetrahedron. If straight lines were used to join each pair of the four clusters, the postulated projection represented in Figure 1 would be reproduced. In other words, the model is consistent with the projections of a cluster pattern, but these projections define only limited regions of this figure.

The fact that this pattern limits the complexity of each test to one accounts for this and also defines the particular regions as regions of complexity one.

Viewed from the point of view of the primary factor pattern, these regions of complexity one must be the primary axes. Further, the primary factor pattern postulates that each primary axis is a vector, rather than say, a plane or a solid, that is coordinate with the common origin of the test vectors. It was seen above that such distinctive vectors are located in the m-space hypertetrahedron by the intersections of the various combinations of m - 1 of the m coordinate hyperplanes. It was also shown that such vectors project as points that form the apexes of the projection. Equating the regions of complexity one with the primary axes is therefore consistent. It follows that any test vector that lies in a region of complexity one

5. See Holzinger and Harman, op. cit., pp. 106-108, for a summary of types of preferred patterns. I have taken the liberty of generalizing their uni-factor pattern to the oblique case and calling it a cluster pattern. I also have split the multiple-factor pattern into more than one type, depending upon the extent of the overlap of the group factors, and used numerals to designate these. Thurstone's criteria for simple structure, op. cit., pp. 334-335, may also be stated in terms of maximum complexity.

6. Two cases deserve consideration. One is the case of an initial orthogonal solution, such as the centroid. The other is the case of an initial oblique solution. In order to simplify the treatment, only the former is discussed here. The principles, however, are general and apply to the extended-vectors representation of an initial oblique solution as well. This fact leads to an interesting variant of Holzinger's simple method of factor analysis; discussion of it will be reserved for a subsequent paper.

7. Only a schematic representation is used in this paper. In preparing the diagrams, however, numerical entries were substituted for the plus signs, the correlation matrix calculated, and the centroids extracted. The primary factor pattern, P, and the intercorrelations of the primaries, $\Phi$, yield R, as follows:

$$P \Phi P' = R$$

See Holzinger and Harman, op. cit., Chapter XI and Appendix B.11.

lies in the intersection of m - 1 of the coordinate hyperplanes of the m-space hypertetrahedron. If it does, then it necessarily has zero coordinates with respect to the axes formed by the other intersections of the various combinations of m hyperplanes taken m - 1 at a time, i.e., with respect to the other primary axes. Viewed in Thurstone's manner, such a test vector is seen to have zero correlations with the normals to the m - 1 hyperplanes at whose intersections it lies. Consequently, it has a non-zero correlation with a normal to only one hyperplane--the one not included in the group of m - 1.

The regions of complexity two are also of considerable interest. A limited 2-multiple-factor pattern will be used to illustrate these. The limitation is that only four of the six possible combinations of complexity two will be included, thus:

| | $P_1$ | $P_2$ | $P_3$ | $P_4$ |
|---|---|---|---|---|
| A | + | 0 | 0 | 0 |
| 1 | + | + | 0 | 0 |
| B | 0 | + | 0 | 0 |
| 2 | 0 | + | + | 0 |
| C | 0 | 0 | + | 0 |
| 3 | 0 | 0 | + | + |
| D | 0 | 0 | 0 | + |
| 4 | + | 0 | 0 | + |

The three projections of this pattern are given in Figure 3. Straight lines have been drawn through the variables indicated by numerals and joining pairs of clusters. These lines pass through all points representing the four given combinations of complexity two. Again, the three plane representations, each a rotation of another through one dimension, are evident. Though incomplete, the projection of the hypertetrahedron is clearly suggested, and we may conclude that the postulated model is consistent with this type of factor pattern. In order to complete the projection it is necessary only to add variables representing the remaining two combinations of complexity two. If this is done, their projections lie between A and C and between B and D on each diagram; these then locate the remaining two straight lines needed to complete the figure.

A region of complexity two constitutes a plane in the m-space hypertetrahedron that is projected as a line. It is bounded, so to speak, by two regions of complexity one, or two of the vectors that are formed by the intersection of m - 1 of the coordinate hyperplanes. In other words, it is a plane that spans two of the primary axes. Test vectors lying in such a plane must have zero coordinates with respect to all primary axes except those two. Since such a plane has been defined above as the intersection of m - 2 hyperplanes, a test vector lying in this plane is therefore uncorrelated with the normals of m - 2 of the coordinate hyperplanes. In general, a test vector lying in a region formed by the intersection of m - x of the coordinate hyperplanes of the m-space hypertetrahedron has a complexity of x, i.e., its primary pattern will have non-zero coordinates with respect to x of the primary axes. A variable of complexity m is therefore a non-distinctive variable (for that factor pattern) since its projection cannot contribute to the delineation of any hyperplane.

Figure 4 gives a schematic representation of all the regions defined by a four-factor pattern in which the maximum complexity of any variable is three. This would be, in the notation used previously, a 3-multiple-factor pattern. A, B, C, and D are each of complexity one, and 1, 2, 3, and 4 are of complexity two, as before. The remaining combinations are:

| | $P_1$ | $P_2$ | $P_3$ | $P_4$ |
|---|---|---|---|---|
| 5 | + | 0 | + | 0 |
| 6 | 0 | + | 0 | + |
| w | + | + | + | 0 |
| x | + | + | 0 | + |
| y | + | 0 | + | + |
| z | 0 | + | + | + |

In combination with A and C, 5 defines a straight line on the projection. This line is the region of complexity two, $P_1$ - $P_3$, and the line joining B, 6, and D is the region of complexity two, $P_2$ - $P_4$. It should be clear that any number of variables might lie along and consequently define such lines in the projection of actual problems. The four regions of complexity three are indicated by the four triangles formed by the six straight lines. Thus the lines forming the triangle ABC define the region w, of complexity three, on each diagram. Again, many variables might lie in such regions in actual problems. Variables of complexity m (four, in this case) have been called non-distinctive variables, since their projections do not define any of these distinctive regions. They may, for example, lie either within or without the four-sided figure, ABCD. In the latter case they have at least one negative pattern value.

The five-factor problem may be interpreted in a similar fashion. In the extended-vector representation, five points define the regions of complexity one, and ten straight lines joining the various pairs of these points define regions of complexity two. The regions of complexity three are seen to be triangles in the plane representation, and the regions of complexity four, four-sided plane figures. For an illustration, the reader should examine Thurstone's fictitious five-factor problem that he used to illustrate his conception of the use of the extended-vectors representation in solving the rotation problem.[8] The regions of complexity one, two, three, and four for this hypertetrahedron can be recognized readily. If a factor pattern similar to those discussed in this paper fits a given six-factor problem, then the projection of the 6-space hypertetrahedron should be consistent with the configuration of extended test vectors on the various diagrams. In theory, at least, m may be any number; the practical problem of recognizing these projections when m is large is another matter.

## Applications and Further Problems

The major point of this paper is that a particular type of configuration can be shown to be a satisfactory geometric model for certain types of primary factor pattern. Such a postulate should make possible several developments and applications. For one, this postulate should clarify considerably the problem of "what to look for" in attempting a primary solution of a given matrix, R. A factor pattern may be regarded as an hypothesis to be tested by the empirical data of a correlation matrix. If a cluster pattern or multiple-factor pattern of maximum complexity less than m is a good fit for the empirical data, then the set of diagrams of the extended-vectors representation of the test configuration will reveal this by their consistent correspondence with the projection of the m-space hypertetrahedron. For a configuration of test vectors made up of exactly m distinct clusters, the choice of a new coordinate system--which constitutes the rotation problem--can be accomplished by taking the m points defined by the projected clusters, plus the origin as the m + 1 points of the simplex of reference, or new coordinate system.[9] The model is here characterized by its correspondence with the configuration of test vectors with respect to regions of complexity one only.

Other problems are less simple. For them, the choice of a new coordinate system may be seen as posing this problem of correspondence with respect to regions of complexity greater than one. If the factor pattern is to be made up only of positive or zero coefficients, then it is evident from this development that the extended-vectors representation of the test configuration must be bounded by regions of complexity two, whose intersections determine the regions of complexity one. Further, the regions of complexity three must also be bounded by regions of complexity two, as must regions of complexity four, five, etc. This view, then, lays great stress on identifying the test vectors that lie in the planes that span the primary axes.[10] Not all students will judge this to be a desirable emphasis. Some, for example, will prefer to take certain variables or clusters of variables as factorially "pure" and then analyze all the variables in terms of these, regardless of the maximum complexity or the signs of the coefficients this gives to the pattern. Both approaches are undoubtedly useful. If one chooses the former approach, however, then this postulate of the m-space hypertetrahedron and its projection should offer some help in the identification of the new coordinate system from the set of extended-vectors diagrams.

This postulate also provides a needed geometric basis for the calculation procedure of direct rotation to primary structure. This procedure has been given in detail previously.[11] Briefly, if the plane representation of the projection of an m-dimensional hypertetrahedron is consistent with the empirical data given as the extended-vectors representation of the centroids, then it should be possible to identify the m regions of complexity one. Each of these is represented by points whose coordinates are consistent on the various diagrams. These coordinates are direction numbers which, when normalized, give the transformation matrix, T, that transforms the centroids directly into the primary structure. Both the intercorrelations of the primary axes and the pri-

8. Thurstone, op. cit., pp. 232-236. Only four of the six diagrams are given, and therefore the remaining two should be constructed. It is now painfully clear that some of my earlier statements regarding the "proper" appearance of these projections were in error. See Chester W. Harris, "Direct Rotation to Primary Structure," Journal of Educational Psychology, XXXIX (December 1948), pp. 449-468.

9. See Sommerville, op. cit., p. 55.

10. This point emphasizes the difference between Thurstone's view of the extended-vectors representation (op. cit., Chapter XI) and mine.

11. Harris, op. cit.

mary pattern may be calculated from these data. The obvious advantage of the procedure is that only a single rotation is made. The geometric formulation given in this paper defines these regions of complexity one as the projections of the primary axes. It also defines the primary axes as vectors formed by the intersection of m - 1 hyperplanes. The procedure of taking the coordinates of the regions of complexity one as the direction numbers of the primary axes thus follows logically.

An important direction in which inquiry might turn is that of the types of figure that constitute the models for other types of factor pattern. It is well known that the factor solution is indeterminate and that one pattern may be transformed into another within the restrictions of certain boundary conditions. It is not known, however, how many different types of factor patterns are distinctive in the geometric sense of demanding different types of figure as the model. For example, the factor pattern that consists of a general factor plus bipolar factors might be studied in this fashion. One suspects that some such patterns are not uniquely bipolar, since their projections would be consistent with the projection of the hypertetrahedron. On the other hand, other bipolar patterns may be unique in the sense that a different model is required. This line of inquiry might lead eventually to the definition of distinctive types of correlation matrices in terms of geometric models.

Another avenue of inquiry may lead to an analytic solution in terms of primary factors. Certain suggestions grow out of the development given here. Any region of complexity two appears as a straight line, bounded by two points of complexity one, on the extended-vectors representation. Further, this region is a straight line on all the diagrams, and all test vectors that project on such a line maintain the same position relative to each other and to the two boundaries on all diagrams. Assume that it is possible to specify, a priori, one of the regions of complexity one. This can often be provided for in designing an actual factor study by including a cluster of tests that are known to be relatively "pure" factorially. This region of complexity one may then be made the origin for the extended-vectors matrix; all that is necessary is to subtract from each column the coordinate of that specified point. It then follows that all variables that consistently lie on a line with that point will have a constant slope for each of the $\frac{(m-1)(m-2)}{2}$ pairs of columns. Assume a five-factor problem, with $w_j$, $x_j$, $y_j$, and $z_j$ as the non-zero extended-vectors coordinates when referred to the specified region of complexity one as origin. The six slopes, for any variable j, are:

$$\frac{w_j}{x_j}, \frac{w_j}{y_j}, \frac{w_j}{z_j}, \frac{x_j}{y_j}, \frac{x_j}{z_j} \text{ and } \frac{y_j}{z_j}$$

All variables defining a region of complexity two must meet the criterion that the sum of these six slopes equals a constant. This can be written:

$$K = a_j + (a_j + 1)b_j + \left[(a_j + 1)b_j + 1\right]c_j$$

where $a_j = \frac{w_j}{x_j}$, $b_j = \frac{x_j}{y_j}$, and $c_j = \frac{y_j}{z_j}$. Such a constant can be defined for m of any size. When expressed in the form given here, it will always have exactly m - 2 terms. Such a criterion, K, will identify all groups of variables that lie in straight lines, on all the diagrams, with the point taken as the arbitrary origin or region of complexity one. In an "ideal" problem, such a procedure would identify m - 1 of the groups of variables whose projections lie in these regions of complexity two.

Variables lying in a region of complexity three must be characterized by a slope, with respect to any pair of axes of the plane representation, that consistently lies between limits given by the slope of variables of complexity two that bound the region of complexity three. A method of sorting variables into those that may and those that may not lie in a given region of complexity three is therefore given. The principle of such a sorting process may be extended to regions of complexity of m - 1. Sorting for variables that lie in regions of complexity three should also suggest one or more groups of variables that, since they may not lie in a region of complexity three, may determine one of the unknown regions of complexity two. Such an hypothesis would then lead to solving the equations of the lines delineated by the groups of variables of constant slope for their consistent points of intersection. In this manner, an analytic solution might be built up. This sketch of the procedure merely suggests possibilities. A rigorous treatment and careful attention to the problem of a satisfactory measure of sampling fluctuation of the criterion K and of the slopes making up this criterion are needed in order to provide a sound statistical basis for such a procedure. Further, these present suggestions are all limited by the requirement that one region of complexity one has been specified a priori.

It seems plausible that the reasoning given in this paper should lead eventually to the formulation of certain conditions that must hold for R if a primary factor pattern of the

type illustrated provides an adequate fit. The determination of these conditions may then lead to a primary solution that is derived directly from R, rather than from an intermediate factor matrix. At present, it appears that these conditions would have to be expressed as relatively complicated relationships in which the unknown communalities would play an important role. For example, a necessary condition for test vectors that lie in a plane is that the angular distance, expressed in radians, e.g., between the two variables lying farthest apart is exactly equal to the sum of the angular distances between each pair of variables taken in sequence with respect to their position. The inclusion of any variable outside this plane in the group would destroy this relation, since the great circle is the shortest path between two points on a sphere. Angular degrees of separation of test vectors can be transformed into cosines of these angles, and the cosines into correlation coefficients divided by the respective lengths of the pair of vectors.[12] Thus, such a condition could be written in terms of the correlation matrix entries. Further exploration would be needed in order to determine the additional conditions that yield the $\frac{(m)(m-1)}{2}$ planes that intersect in m common points. Exploration of the possibility of writing these conditions and then of making use of them to secure a direct solution would seem to be warranted, however.

12. This assumes that m is to be taken as the minimum rank of R.

Figure 1




Figure 2

3a




3b




3c

Figure 3




4a




4b




4c

Figure 4

# THE DEVELOPMENT OF A TECHNIQUE FOR THE MEASUREMENT OF SOCIAL-EMOTIONAL CLIMATE IN CLASSROOMS

JOHN WITHALL*

University of Chicago

Rationale.--In the study[1] proper a brief overview of current theories of learning from the associationist and field-theorist point of view precedes a statement of the major concepts, gleaned largely from the field-theorists, that guide the study. Certain postulates regarding the motivational factors in personality and regarding learning conditions are proposed. The primary motivational force of human behavior is postulated to be a drive toward self-actualization. This drive is said to be influenced by:

1. Need for self-consistency
2. Interaction in terms of an internal frame of reference
3. Self-directive behavior
4. Achievement of personal significance and private meanings in a social milieu.

It is postulated that learning (changes in behavior) is most likely to occur when experiences are both

1. meaningful to the learner, that is, are perceived by the learner as pertinent to his needs and purposes, are consistent with his personality organization, and are associated with self-directive behavior; and

2. occur in a non-threatening situation, that is, the learner is free from a sense of personal threat, interacts with others in a wholesome social milieu, and is helped to evaluate himself on the basis of objective criteria.

Since condition number 1 is postulated to be dependent on condition number 2, some knowledge of the psychological atmosphere represented by number 2 is necessary. Little objective evidence exists regarding psychological atmospheres either in learning or other situations. It was decided, therefore, to attempt to develop a technique to measure social-emotional climate in the classroom through a categorization of teacher-statements. It seems reasonable to assume that the teacher's behavior influences the conditions of learning since she is placed in the classroom by society to manipulate the conditions so as to facilitate learning.

Assumptions and hypothesis.--The basic assumptions of the study are (1) that the social-emotional climate is a group phenomenon; (2) that the teacher's behavior is the most important single factor in creating climate in the classroom, and (3) that the teacher's verbal behavior is a representative sample of her total behavior.

The hypothesis to be tested is that by means of a categorization of teacher-statements a valid and reliable index of social-emotional climate can be obtained.

## Work Previously Done in Area of Climate

Some work has been done in the area of climate, notably by Ronald Lippitt at Iowa University and Harold H. Anderson at Michigan State College.

* The author is indebted to several members of the faculty of the Department of Education at the University of Chicago for their counsel and constructive criticisms during the carrying out of the study reported here. He is particularly grateful to both Dr. Herbert A. Thelen of the Department of Education and the Department of Psychology, and to Dr. Carl R. Rogers of the Department of Psychology, University of Chicago, for their facilitation of the experimenter's efforts to make the study as scientifically sound and as theoretically fruitful as possible.

1. John Withall, The Development of a Technique for the Measurement of Social-Emotional Climate in Classroom. Unpublished Ph.D. dissertation, Department of Education, University of Chicago, 1948.

### Lippitt's Study

Lippitt's work in social climate is best represented by his doctoral study.[2] In that study he organized four clubs of five boys each and gave each club successive experiences with an "autocratic" and "democratic" leader during three consecutive six-week periods. Several leaders were used to head each of the clubs. The leaders were required to employ different leadership styles with each successive group. The leadership styles were implemented in accord with certain specific criteria drafted to guide the club leaders. Records of social interaction between group members and leader, stenographic records of conversation in each club, analysis of activity subgroupings, and a running account of psychologically interesting interaction in each group were among data collected by observers of each club session.

The major conclusions of Lippitt's study were:

1. That different leadership styles produced different social climates and resulted in different group and individual behaviors;
2. That conversation categories differentiated leader-behavior techniques more adequately than social-behavior categories;
3. That autocratic leadership elicited either an aggressive rebelliousness towards the leader or an apathetic submission to the leader;
4. That leadership style was the primary factor in producing climatological differences and that club personnel was of secondary importance.

Lippitt's work represents one of the earliest and most significant attempts to observe and control the climate variable in a group situation. His findings regarding the value of categorizing verbal behavior as a means of assessing the quality of group life provides a sound basis for the methodology of this study which utilizes a categorization of teachers' verbal behaviors as its major technique.

### Anderson's Work

Harold H. Anderson, Joseph E. Brewer and others have conducted at Michigan State College investigations[3] into the influence of teachers' classroom personalities on children's behavior, particularly at the primary and elementary school levels. In order to obtain objective measurements of teachers' classroom personalities and concomitant children-behavior, Anderson and Brewer developed twenty-six teacher-behavior categories and twenty-nine children behavior categories by which both teacher and pupil verbal and non-verbal behavior might be categorized. Anderson divided teacher-behaviors into two main kinds -- Integrative Teacher Behavior and Dominative Teacher Behavior. Integrative behavior was that which expanded the children's opportunities for self-directive and cooperative behavior with the teacher and their peers; dominative behavior tended to restrict children's activities and to lead to distracted, aggressive, noncooperative conduct. Anderson demonstrated that children's behaviors were consistent with the kind of personality the teacher displayed in the classroom.

Anderson's studies bring out evidence that is highly pertinent to the hypothesis that the main direction of influence in the classroom is from the teacher to the pupil. He has demonstrated too, that reliable patterns of teacher and pupil-behavior can be obtained in the classroom through categorizations of their overt behaviors.

### Development of Climate Index

The concept of climate or psychological atmosphere has been used by others in the area of psychology and education besides Lippitt and Anderson. Lewin, Prescott and Rogers, for example, have made considerable use of the concept. However, no clear-cut definition of the concept can be cited and for the purpose of more effective communication and clearer understanding of the notion as used here, a definition of the term "social-emotional climate" is offered.

Definition of social-emotional climate.--Climate is considered in this study to represent the emotional tone which is a concomitant of interpersonal interaction. It is a general emotional factor which appears to be present in interactions occurring between individuals in face to face groups. It seems to have some relationship to the degree of acceptance expressed by members of a group regarding each other's needs or goals. Operationally defined it is

2. Ronald Lippitt, An Analysis of Group Reaction to Three Types of Experimentally Created Social Climate. Unpublished Ph.D. dissertation, University of Iowa, 1940.

3. Harold H. Anderson, et al. Studies of Teachers' Classroom Personalities, Applied Psychology Monographs, Nos. 6, 8, and 11 (Stanford, California: Stanford University Press).

considered to influence: (1) the inner private world of each individual; (2) the esprit de corps of a group; (3) the sense of meaningfulness of group and individual goals and activities; (4) the objectivity with which a problem is attacked, and (5) the kind and extent of interpersonal interaction in a group.

Procedure for identifying categories.--Proceeding from our assumption that climate is largely determined by the teacher's behavior, we undertook an analysis of the teacher's verbal-behaviors contained in sound recordings of regular class sessions. A teacher's verbal-behavior is assumed to represent adequately her total behavior. Guided by the postulates regarding individual motivation and conditions of learning, teacher-statements were analyzed in order to ascertain whether the teacher was utilizing behaviors likely to create the postulated conditions for learning. Individual teacher-statements tended to fall into about twenty-five types of responses, e.g.:

1. Reproof statements made with the apparent intent of halting the pupil's present behavior and modifying future behavior;
2. Questions seeking further information from pupil about problem;
3. Statement offering simple administrative information, e.g., about place of next class meeting;
4. Questions containing reproof and a plea for cooperation;
5. Statement approving pupil's behavior and commending him;
6. Statement urging a particular course of action on the learner; and so on.

These twenty-five kinds of responses were soon found to overlap and to be not mutually exclusive. They were reduced to thirteen and finally to seven categories. These categories seemed to encompass all the types of statements which teachers utilized in classrooms. The categories are:

1. Learner-supportive statements that have the intent of reassuring or commending the pupil.
2. Acceptant and clarifying statements having an intent to convey to the pupil the feeling that he was understood and help him elucidate his ideas and feelings.
3. Problem-structuring statements or questions which proffer information or raise questions about the problem in an objective manner with intent to facilitate learner's problem-solving.
4. Neutral statements which comprise polite formalities, administrative comments, verbatim repetition of something that has already been said. No intent inferrable.
5. Directive or hortative statements with intent to have pupil follow a recommended course of action.
6. Reproving or deprecating remarks intended to deter pupil from continued indulgence in present "unacceptable" behavior.
7. Teacher self-supporting remarks intended to sustain or justify the teacher's position or course of action.

It seemed that more than one continuum was identifiable in the seven categories; for example, a continuum from problem-centeredness to person-centeredness might be discerned, a continuum from objectivity to subjectivity or a continuum from learner-centeredness to teacher-centeredness. The latter continuum was accepted as useful in applying the categories to teachers' verbal behaviors. Categories 1, 2, and 3 were said to be learner-centered and categories 5, 6, and 7 were said to be teacher-centered. The neutral category had no influence on either bloc. Certain conventions were suggested for interpreting patterns of verbal behavior; for instance, if the proportion of statements falling into one or more of the first three categories outweighed the proportion falling into one or more of the last three categories, the teacher was said to be learner-centered. If the proportion were reversed, the teacher was said to be teacher-centered. Similarly, if the proportion of statements falling into category 3 outweighed the proportions falling into either categories 1 and 2 combined or categories 5, 6, and 7 combined, then the teacher was said to be more problem-centered than learner or teacher-centered. If the largest proportion fell into the combination of either categories 1 and 2 or categories 5, 6, and 7, then the teacher was said to be more highly learner-centered or more self-centered as the case might be. Once the seven categories had been identified and procedures developed for facilitating their application to data, the next step was to ascertain the objectivity, reliability, and validity of the technique.

## Determining the Objectivity, Reliability, and Validity of the Climate Index

### Objective Study

The objectivity of the instrument was ascertained by having four trained judges apply the index independently to typescripts 1A, 2A, and 3A comprising 68, 71, and 45 teacher-statements respectively. The percentage of agreement of each judge with the researcher was computed. The percentage of agreement was computed by the formula:

$$\text{Percentage of agreement} = \frac{\text{Number of identical statements which researcher and judge place in a given category}}{\text{Number of statements which researcher places in same category}} \times 100$$

The mean percentage of agreement between the judges and researcher on the three typescripts was:

| | |
|---|---|
| Judge 1 | 64 percent |
| Judge 2 | 66 percent |
| Judge 3 | 56 percent |
| Judge 4 | 75 percent |

The highest mean percentage of agreement of the four judges with the researcher on the three typescripts was 78 percent on category 1 and the lowest 53 percent on category 7. The grand mean percentage of agreement for four judges with the researcher on the three typescripts was 65 percent.

A further measurement of the index's objectivity was made by computing tetrachoric correlations between the categorization of each of the five judges (the researcher included) on the three typescripts. Two hundred ten $r^{tets}$ were computed to ascertain the degree of association between each of the judges with one another on each category in turn.

The range of correlation and median $r^{tet}$ correlation of five judges' was:

| Typescript | Range | Md. |
|---|---|---|
| 1A | -.09 to .95+ | .84 |
| 2A | .20 to .95+ | .76 |
| 3A | .36 to .95+ | .93 |

### Reliability Study

The size of an adequate sample was empirically determined by the classic procedure of adding further data to ascertain at what point addition of further data caused little or no change in the obtained pattern of statements. Successive blocs of fifty statements from the typescripts of one class session were added to each other and the fluctuation in pattern of statements falling into the seven categories noted. (Table I)

Standard errors were computed to ascertain whether the differences between the successive percentages in each category and the mean were significant, that is, were three times the standard error. No significant differences were found except with regard to the percentage of category 3 statements based on the first fifty and first one hundred statements respectively, and for the first fifty statements of category 4. The other 32 differences were not reliable.

It was concluded that 200 statements would offer an adequate sample of a given teacher's statements since the statements processed represented a class session drawn at random from several class sessions. The class from which the sample was drawn was, as regards size of group, resources, classroom arrangement, et cetera, in no way peculiar or biased.

A further check on the consistency of the index was made by comparing the day to day variations of the pattern of statements of three teachers, A, B, and C.

The chi-square test was used to check the hypothesis that no significant differences occurred from day to day. It was not possible to reject the hypothesis insofar as teachers B and C were concerned. The P's for these samples were .14 and .42 respectively. In the case of teacher A, however, a P of .02 was obtained which suggested that other than chance factors were operative. An "explanation" for this low P might be found in the fact that sample 7 alone contributed more than half the total value of chi-square and was a deviate session in that the teacher, contrary to her usual pattern of behavior was compelling the class to carry out a project to which they objected bitterly. The whole tone of the class session was greatly at variance with the usual tenor of this teacher's session. If this sample 7 is omitted from the series, the P obtained for the rest is .54 which does not permit of the rejections of the hypothesis and permits us to say the differences are insignificant.

### Validation Procedures

The validity study has employed four procedures to validate the technique developed.

1. We have used H. H. Anderson's Teacher Behavior Categories[4] as a criterion instrument that has been applied to the same data as the climate index.

TABLE I

PERCENTAGE OF INCREMENT OR DECREMENT IN THE PROPORTION OF STATEMENTS FALLING INTO EACH OF SEVEN CATEGORIES AS ADDITIONAL BLOCS OF STATEMENTS ARE ADDED

| | Categories: | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| | % | % | % | % | % | % | % |
| First fifty statements | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Fifty added to 50 | 0.0 | 0.0 | 2.0 | -7.0 | 3.0 | -1.0 | 3.0 |
| Fifty added to 100 | 2.7 | 0.7 | 3.3 | -3.7 | -1.7 | -0.3 | -1.0 |
| Fifty added to 150 | -0.2 | -0.7 | 3.7 | -0.8 | -2.7 | -0.2 | 0.5 |
| Fifty added to 200 | -0.1 | 0.0 | 1.0 | 0.4 | -1.0 | -0.1 | -0.2 |
| Twenty-one added to 250 | -1.6 | -0.3 | 1.0 | 0.0 | 0.2 | 0.4 | 0.3 |

TABLE II

I.D. RATIOS AND CLIMATE INDEX RATIOS ON TYPESCRIPTS 1A, 2A, and 3A

| Typescript | I.D. ratio | Ratio of $\frac{\text{learner-centered statements}}{\text{teacher-centered statements}}$ |
|---|---|---|
| 1A | 6.3 | 2.4 |
| 2A | 1.6 | 2.2 |
| 3A | 0.4 | 0.2 |

2. We have related pupil's recorded comments and feelings to the picture of climate in the classroom situation obtained by application of the climate index to the teacher's statements.

3. We have predicted from results obtained on typescripts the ratings that independent judges would make of teacher-behaviors in terms of a Teacher-Characteristics Rating Scale.

4. We have related the description of a classroom situation from three frames of reference--a categorizer using objective criteria, judges' ratings of the teacher's verbal behavior in a "live" situation and pupils registering their negative or positive feelings.

Use of Anderson's Teacher-Behavior Categories.--Anderson has developed Teacher-Behavior Categories by which to describe Teacher's Classroom Personalities. The categories distinguish between Integrative and Dominative teacher-behavior and are the basis for an I.D. ratio[5] which represents the ratio of approving, cooperative, integrative behavior by the teacher. It can be shown that the I.D. ratio in many respects is similar to a proportion of learner-centered to teacher-centered verbal behaviors. I.D. ratios and learner-centered to teacher-centered ratios are comparable since in each instance a ratio greater than 1.00 denotes a greater use of integrative, acceptant and facilitating behaviors over dominative, unacceptant and limiting behaviors by the teacher.

Anderson's categories and the climate index categories were applied to the same data with the results found in Table II.

Analyses by both Anderson's and the climate index criteria appear to concur that a greater proportion of teacher-behaviors in the situation represented by typescript 1A and 2A were integrative and acceptant whereas a greater proportion of teacher-behaviors in 3A were dominative and rejecting.

Pupils' evaluations.--The pupils' evaluation of the teacher's methods and of the classroom situation in general is admissible evidence which may sustain or contradict the picture of climate obtained in a given classroom. We have such pupil evaluations regarding teacher A in class W. We also have climate index indications of the major focus of concern of the teacher derived from categorizing typescripts 1A and 2A as well as a mean pattern obtained from 8 random samples drawn from the same teacher's class sessions. The climate index analysis shows that, on the whole, in order of importance the teacher's concern is first with herself, second with the learner, and third with the problem. (Table III)

There is some evidence here which suggests that the teacher tended to place more emhasis on interpersonal relationships than on structuring the objective problem in the learning situation. The concensus of opinion on the part of the 25 pupils who filled out evaluation sheets was ambivalent and can be best summarized as follows:

Approving

Teacher was cooperative
Teacher was unprejudiced and impartial
I got new ideas
Teacher helped me

Disapproving

I wasted considerable time
Teacher ought to have guided and controlled class more
I was bored
Teaching method was unsatisfactory

The picture obtained both from the climate index and from the pupils' evaluations tended to concur in depicting a teacher who evidenced more concern with personality factors in the learning situation than with the objective problem elements.

The comparison of these data afford another crude indication that the climate index categorization tends to represent with some accuracy the actual atmosphere obtaining in a classroom, at least, as seen through the eyes of the pupils.

Teacher-Characteristics Scale.--Four judges who had not used the climate index and who had no knowledge of the criteria contained in the index were asked to scale typescripts 1A, 2A, and 3A according to two dichotomous lists of teacher-behaviors which were labelled "A" and "C" on a 9-point scale. The midpoint was marked "B".

Point "A" was defined by such teacher-behaviors as:

1. Raising objectivite questions about procedures and purposes of learner

2. Evaluating pupil-achievement in terms of stated, objective criteria

4. Anderson et al., op. cit., pp. 22-26.

5. Anderson et al., op. cit., pp. 41-42.

TABLE III

PERCENTAGE OF TEACHER A'S STATEMENTS FALLING INTO EACH OF THREE AREAS OF CONCERN

| Areas of Concern | Typescript 1A Percent* | Typescript 2A Percent* | Mean for Eight Random Samples Percent* |
|---|---|---|---|
| Sustain herself (Categories 5, 6, and 7) | 28.0 | 77.9 | 37.4 |
| Helping the learner (Categories 1 and 2) | 44.0 | 13.3 | 23.4 |
| Developing the problem (Category 3) | 22.1 | 4.4 | 28.8 |

*The remaining percentage in each instance is accounted for by the neutral category, number 4.

TABLE IV

COMPARISON OF WEIGHTED RATINGS OF TEACHER-CHARACTERISTICS RATING SCALE WITH CLIMATE INDEX RATIO OF LEARNER-CENTERED TO TEACHER-CENTERED STATEMENTS ON TYPESCRIPTS 1A, 2A, and 3A

| Typescripts | Weighted Ratings | | | | Ratio of Learner-centered statements / Teacher-centered statements |
|---|---|---|---|---|---|
| | Judge 5 | Judge 6 | Judge 7 | Judge 8 | |
| 1A | 0.0 | 1.3 | 3.0 | 3.7 | 2.4 |
| 2A | 3.0 | 1.0 | 0.8 | 3.0 | 2.2 |
| 3A | -2.5 | 0.5 | -2.0 | -3.5 | 0.2 |

3. Helping learners to identify a wide choice of problem-solving activities, etc.

Point 'C" was defined by such teacher-behaviors as:

1. Limiting choices of behavior of the learner

2. Evaluating the learner's achievement on the basis of the teacher's unstated, subjective criteria

3. Setting the pace of problem-solving behaviors for the learner, etc.

Ten such behaviors for "A" and "C" were listed. The training given the judges was comparatively brief--a forty minute session--and then they were asked to apply the scale to typescripts 1A, 2A, and 3A. From the categorization of statements on the basis of the climate index criteria, it was predicted that the judges would place their ratings on the 9-point scale between "A" and "B" for typescripts 1A and 2A, and between "B" and "C" for typescript 3A. The results were that the ratings were as predicted for typescripts 1A and 2A except that on 1A Judge 5 placed his rating exactly at "B" and were as predicted for 3A except that Judge 6 placed his rating on the "positive" side of "B" instead of on the "negative" side as predicted. Positive weights were given to the points from "B" to "A" with "B" marked zero and "A" marked plus 4; negative weights were given the steps from "B" to "C" being weighted minus 4. The resultant weighted ratings are represented in Table IV.

Description of class situation from three frames of reference.--A further validation procedure involved the relating of data describing an experimental classroom situation from three frames of reference:

1. The objective frame of a categorizer using the criteria of the climate index

2. The frame of observers rating teacher-statements in situ with the aid of illustrative teacher-behaviors on a dichotomous scale

3. The frame of the pupils in the classroom

The experimental session was forty-five minutes long. In the first part of the period, the teacher sought by means of his verbalizations to create a learner-centered climate and in the last part of the period a teacher-centered climate.

The data that are reported here refer to session Number 2 of Group "X" comprising ten pupils. In the categorization of teacher-statements by means of the climate index, it was found that of the eighty-three statements in the learner-centered part of the session:

48.2 percent were learner-centered
12.0 percent were teacher-centered
39.8 percent were neutral

Of the thirty-three neutral statements comprising 39.8 percent neutral statements, twenty-one were simple verbalizations of pupils' names.

Of the 119 statements in the teacher-centered part of the session:

31.9 percent were learner-centered
55.5 percent were teacher-centered
12.6 percent were neutral

Of the fifteen neutral statements comprising the 12.6 percent, four were simple verbalizations of pupils' names.

There is an obvious difference between the learner-centered and teacher-centered parts of the class session as regards the percentages of teacher-centered statements. The difference, however, between the proportions of learner-centered statements is not large quantitatively but differs considerably qualitatively. For example, the percentage of category 1 remarks is nil and the bulk of the learner-centered statements falls into category 2 during the learner-centered portion of the lesson, whereas in the teacher-centered half of the lesson 24.4 percent of the total 31.9 percent of learner-centered statements fall into category 1. These category 1 statements, described as learner-supportive, were mostly superficially commendatory remarks such as "Right!" "Good!" and "Very good!"

The categorizations by means of the climate index show that there was a preponderance of learner-centered remarks over teacher-centered remarks in the first portion of the session and a preponderance of teacher-centered over learner-centered remarks in the second portion of the session.

During the class session proper, four judges who were trained therapists rated the teacher's statements in terms of illustrative criteria drawn from the Teacher-Characteristics Rating Scale. One end of the scale was designated "A", the opposite end "C" and the midpoint "B". If a judge rated a statement as falling between "A" and "B", it was interpreted as being learner-centered; if he rated between

TABLE V

MEAN PERCENTAGES OF JUDGES' RATINGS OF TEACHER-STATEMENTS USED DURING LEARNER-CENTERED PART OF SESSION

| Mean Percent | Kind of Statement | Range of 4 Judges' Percentages |
|---|---|---|
| 63.8 | learner-centered | 45.0* to 77.7 |
| 18.6 | neutral | 10.0 to 22.3 |
| 17.6 | teacher-centered | 0.0 to 45.0* |

* Judge 4 was responsible for the 45 percent item in both instances. He explained afterwards that his bias had been to scale objective questions which sought information as probing-directive statements; the other judges interpreted them as problem-structuring and therefore as learner-centered comments.

TABLE VI

MEAN PERCENTAGES OF JUDGES' RATINGS OF TEACHER-STATEMENTS USED DURING TEACHER-CENTERED PART OF SESSION

| Mean Percent | Kind of Statement | Range of 4 Judges' Percentages |
|---|---|---|
| 2.9 | learner-centered | 1.5 to 5.7 |
| 6.9 | neutral | 1.5 to 12.8 |
| 90.2 | teacher-centered | 81.5 to 97.0 |

"B" and "C", it was considered to be teacher-centered; if he rated it at "B", it was interpreted as a neutral statement.

In the learner-centered part of the session the mean percentages of the four judges' ratings of teacher-statements are found in Table V, and in the teacher-centered portion of the session the mean percentages of the four judges' ratings are found in Table VI.

The data arrived at from an objective categorization of statements by means of the climate index and that obtained from four judges' ratings of statements in situ are in general agreement.

For the experimental class periods each of the ten pupils of Group "X" was supplied with two push buttons; one labeled plus and the other minus. They were requested, (1) to push the plus button whenever they felt "good" at what the teacher said whether the remark was directed at them or not, and (2) to push the minus button whenever they felt badly at what the teacher said. An orientation session was held with the group prior to the experiment proper to get them acquainted with the equipment and procedure but no training was given them regarding the kinds of statements which were to evoke either negative responses or positive responses.

It was the prediction of the researcher that more positive than negative reactions from the pupils would be registered in the learner-centered portion of the session, and more negative than positive reactions in the teacher-centered part of the session. The recorded pupil reactions are found in Table VII.

The positive reactions slightly predominated during the learner-centered part of the class session and the negative reactions predominated during the teacher-centered part of the class.

These data derived from three frames of reference: (1) a categorizer using climate index criteria, (2) the observers' ratings of statements in situ, and (3) the pupils' "feeling" reactions, on the whole, tend to concur in depicting the first part of the session as probably more learner- and problem-oriented, with a greater possibility of objective attitudes by the teacher and pupils to the learning situation, and the latter part of the lesson as more teacher centered and more conducive to a subjective[6] orientation by the teacher and pupils to the learning situation.

By the procedures outlined in the preceding pages, the climate index was shown to have objectivity, reliability, and validity.

## Application of Climate Index to Excerpts from Five Different Classes

To test the applicability of the index categories to regular classroom sessions, samples were drawn from the sound recordings of five regular classes held in the laboratory classroom. These classes included a 7th grade mathematics class, two 8th grade social science classes, an 8th grade art class, and a 9th grade Latin class. The classes met regularly in the classroom three or four times a week and were conducted by the regular class teacher in her usual manner.

Sampling.--Eight random samples, each seven minutes in length, were drawn from the recorded class sessions of the art, mathematics, and Latin class. In addition, one full class session was taken from each of the three aforementioned classes as well as from the two social science classes. A total of 23 seven-minute excerpts (one sample from the Latin class was unusable) and five full class sessions were available from classes at the secondary school level. These excerpts from the sound records of the several class sessions were typed-up and the seven categories applied to the teacher-statements contained in the protocols.

Patterns of statements for each of the four teachers were obtained and an attempt was made, largely for illustrative purposes, to interpret these patterns. A mean pattern of distribution of statements among the seven categories was obtained for each of teachers A, B, and C from the several excerpts from their respective classes. A pattern of statements for teacher D in classes W and Z, respectively, was also obtained from two full class sessions of the social science classes. (See Table VIII)

Comment regarding teachers' patterns of statements.--Teacher B appears to use a larger proportion of learner-centered statements (categories 1, 2, and 3) than the other three. Teachers A, C, and $D_Z$ seem to use a somewhat similar proportion of teacher-centered remarks; teacher B uses a low proportion of teacher-centered remarks (categories 5, 6, and 7). Teacher D in both sessions uses a slightly larger proportion of problem-structuring remarks (category 3) than the other three teachers and teacher A uses the lowest percent-

6. Greater concern with one's self and one's status in the group than with the objective problem.

TABLE VII

COMPARISON OF NUMBER AND PERCENTAGE OF POSITIVE AND NEGATIVE "FEELING" REACTIONS OF PUPILS RECORDED DURING THE LEARNER-CENTERED AND TEACHER-CENTERED PARTS OF A CLASS SESSION

| "Feelings" | Learner-Centered Portion of Session | | Teacher-Centered Portion of Session | |
|---|---|---|---|---|
| | Number of Reactions | Percent | Number of Reactions | Percent |
| Positive | 54 | 54.0 | 72 | 30.2 |
| Negative | 46 | 46.0 | 166 | 69.8 |
| Total | 100 | 100.0 | 238 | 100.0 |

TABLE VIII

PERCENTAGE OF STATEMENTS LOCATED IN EACH OF THE CLIMATE INDEX CATEGORIES FOR TEACHERS A, B, C, AND D

| | Teacher | | | | |
|---|---|---|---|---|---|
| | A | B | C | $D_W$ | $D_Z$ |
| Category | Mean % | Mean % | Mean % | Mean % | Mean % |
| 1 | 18.2 | 20.3 | 8.6 | 2.5 | 4.7 |
| 2 | 5.2 | 7.8 | 2.8 | 0.0 | 0.9 |
| 3 | 28.8 | 39.7 | 30.7 | 53.0 | 41.1 |
| Total | 52.2 | 67.8 | 42.1 | 55.5 | 46.7 |
| 4 | 10.4 | 22.2 | 20.2 | 21.4 | 16.8 |
| 5 | 14.8 | 6.8 | 23.9 | 13.7 | 20.6 |
| 6 | 9.1 | 1.6 | 8.9 | 6.8 | 10.3 |
| 7 | 13.5 | 1.6 | 4.9 | 2.6 | 5.6 |
| Total | 37.4 | 10.0 | 37.7 | 23.1 | 36.5 |
| Grand Total | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| | N=497 | N=408 | N=525 | N=117 | N=107 |

age of such comments. Teacher B uses the largest proportion of category 1 statements and of category 2 remarks. In the teacher-centered area, teacher $D_2$ and teacher C use the largest proportion of directive (category 5) statements, and teacher B the lowest proportion of such statements. Teacher B uses the lowest proportion of reproving (category 6) remarks and of self-supportive (category 7) statements.

A possible interpretation of B's pattern of statements.--Teacher B would appear to be an individual who would offer verbal support and encouragement to the pupils (category 1) and would attempt, to a lesser extent, to convey to them her understanding of them and of their point of view. She possibly tries to keep the objective problem clearly defined and attempts to keep it the central object of attention and concern (category 3). She appears to utilize a small proportion of hortative statements (category 5). Her negative evaluations of pupil-behavior appear to be minimal (category 6). She appears to feel slight need for indulging in self-supportive and defensive comments. Here is a teacher, we might infer, whose methodology for facilitating learning is (1) to keep the learners well oriented to the objective problem, and (2) to maintain a helpful and understanding attitude towards them. At the same time, she would seem to be able to keep her own needs well in the background.

## Application of Climate Index to a Higher Level of Instruction

To test the applicability of the climate index categories to a level of instruction beyond the secondary school level, typescripts derived from recorded sessions of a graduate seminar and from a counselor-training program were processed.

### Pattern of Statements Derived from a Categorization of Professor E's Verbal Behavior

The protocols containing Professor E's verbalizations are based on two 2-hour seminar meetings of graduate students. These were two consecutive sessions that were recorded and upon which the typescripts were based. (Table IX)

Comments on Professor E's pattern of statements.--A high degree of problem-orientation is evidenced in these two class situations by the large proportion of problem-structuring comments (category 3). Positive evaluations (category 1) are minimal and negative evaluations (category 6) are non-existent. Some verbal expression of acceptance of the students' ideas and feelings is in evidence on the basis of the percentages falling into category 2. The proportion of hortative statements (category 5) is not large. However, some evidence of a need to sustain himself and to be concerned about his status appears from the comparatively large (in relation to the rest of the pattern) proportion of category 7 statements in both sessions.

### Pattern of Statements Derived from Categorization of Instructor F's Verbal Behavior

The protocols of Instructor F's sessions were based on three $1\frac{1}{2}$ hour sessions. They were random selections from a group of eight available typescripts. The paucity of statements by the instructor, 150 in all for the entire three sessions, is both striking and interesting. (Table X)

Comments regarding Instructor F's pattern of statements.--Instructor F uses a large proportion of clarifying and acceptant statements (category 2) in all three sessions. Encouraging and reassuring statements (category 1) are at a very minimum. Slightly less than 1/5 of his remarks are problem-structuring comments. Directive, reproving, and self-supportive statements are rarely used and are nearly all confined to session 1; the fact that the bulk of F's self-concerned statements fell into session 1 may be "explained" by the fact that it was the first meeting of the class. The whole pattern of F's verbal behavior is explicable largely in terms of his adherence to the principles of client-centered psychotherapy. Counselors of that persuasion place great emphasis on the creation of permissive and acceptant atmosphere by the use of clarifying and acceptant statements and other appropriate procedures. Instructor F clearly transferred his counseling orientation into the classroom situation.

It would appear from the analysis of both Professor E's and Instructor F's statements that the climate index categories are applicable to a higher level of instruction other than the secondary school level.

## Summary, Conclusions, and Implications

Summary.--A technique has been developed for assessing the social-emotional climate in a classroom by categorizing teacher-statements contained in typescripts made from sound records of class sessions. The technique has been shown to have objectivity, reliability, and validity.

Conclusions.--

1. Climate can be assessed and described.

2. Several individuals can be trained to

TABLE IX

PERCENTAGES OF PROFESSOR E'S STATEMENTS FALLING INTO EACH OF THE SEVEN CATEGORIES

| Category | Seminar 1 (N=102) Percent | Seminar 2 (N=84) Percent |
|---|---|---|
| 1 | 2.0 | 3.6 |
| 2 | 4.9 | 8.3 |
| 3 | 70.6 | 59.5 |
| Total | 77.5 | 71.4 |
| 4 | 8.8 | 11.9 |
| 5 | 5.9 | 1.2 |
| 6 | 0.0 | 0.0 |
| 7 | 7.8 | 15.5 |
| Total | 13.7 | 16.7 |
| Grand Total | 100.0 | 100.0 |

TABLE X

PERCENTAGE OF INSTRUCTOR F'S STATEMENTS FALLING INTO EACH OF THE SEVEN CATEGORIES

| Category | Session 1 (N=36) Percent | Session 2 (N=62) Percent | Session 3 (N=52) Percent |
|---|---|---|---|
| 1 | 0.0 | 8.0 | 0.0 |
| 2 | 63.9 | 58.1 | 69.2 |
| 3 | 22.2 | 21.0 | 15.4 |
| Total | 86.1 | 87.1 | 84.6 |
| 4 | 0.0 | 11.3 | 15.4 |
| 5 | 2.7 | 0.0 | 0.0 |
| 6 | 5.6 | 1.6 | 0.0 |
| 7 | 5.6 | 0.0 | 0.0 |
| Total | 13.9 | 1.6 | 0.0 |
| Grand Total | 100.0 | 100.0 | 100.0 |

use the criteria of the climate index and achieve an adequate measure of agreement among one another in categorizing statements in typescripts.

3. A valid measure of social-emotional climate of groups is obtainable through a categorization of teacher-statements.

4. Within the limits of behavioral and personality variations of the climate index gives us a consistent pattern of verbal behavior for a given teacher from day to day.

5. Different patterns of verbal behavior used by several teachers can be identified.

6. Statements categorized by the climate index as likely to produce "positive" feelings tend to be similarly categorized by impartial observers and tend to be reacted to with "positive" feelings by the individuals to whom they are addressed.

7. Statements categorized according to the climate index as likely to produce "negative" feelings tend to be similarly categorized by impartial observers and tend to be reacted to with "negative" feelings by the individuals to whom they are addressed.

Some revision of the climate index seems in order as a result of insights arising out of its application to several typescripts and to "live" classroom situations. These improvements should include:

1. Clarification of category 1 in order to distinguish between objective and subjective positive evaluation of pupil behavior; the former may be genuinely learner-supportive; the latter doesn't appear to perform that function; clarification of category 5

2. More rigid definition of statements to be placed in category 4 which should contain perhaps only verbatim repetitions, polite formalities and administrative comments

3. Differentiation between types of problem-structuring statements

4. Considering the possibility of sub-dividing five of the seven categories (not 4 and 7) into "A" and "B" areas. The "A" area would represent little or no element of self-concern on the part of the teacher; the "B" area would represent a considerable element of self-concern and result in tingeing the statement with some affect.

Implications for further research.--Some questions that arise out of this study are:

1. What is the relationship between climate and the quality of the learning that occurs in a classroom?

2. To what extent is the climate in a given classroom a function of the personality of the teacher?

3. To what extent do peer-group relationships influence the classroom climate?

4. Of what value is the climate index to teachers in analyzing their own teaching methods?

5. How applicable is the climate index to the "live" classroom situation?

The study described above represents an attempt to develop a technique for the measurement of social-emotional climate to the end that, ultimately, fuller understanding and control may be achieved of one of the factors hypothesized to influence learning.

## BIBLIOGRAPHY

Allport, Gordon W. "Psychology of Participation," Psychological Review, LII (1945), 117-132.

Anderson, Harold H., and Brewer, Helen M., "Studies of Teachers' Classroom Personalities, I. Dominative and Socially Integrative Behavior of Kindergarten Teachers." Applied Psychology Monograph, No. 6, American Psychological Association (Stanford, California: Stanford University Press, 1945), 157 pp.

Anderson, Harold H., and Brewer, Joseph E., "Studies of Teachers' Classroom Personalities, II. Effects of Teachers' Dominative and Integrative Contacts on Children's Classroom Behavior." Applied Psychology Monograph, No. 8 (Stanford, California: Stanford University Press, 1946), 128 pp.

Anderson, Harold H., and others. "Studies of Teachers' Classroom Personalities, III. Follow-up Studies of the Effects of

Dominative and Integrative Contacts on Children's Behavior." Applied Psychology Monograph, No. 11 (Stanford, California: Stanford University Press, 1946), 156 pp.

Angyal, Andras. Foundations for a Science of Personality (New York: Commonwealth Fund, 1941), 398 pp.

Brotemarkle, R. A. "Clinical Point of View in Education," Training School Bulletin, XLIV (October 1947), 102-110.

Cantor, Nathaniel. Dynamics of Learning (Buffalo, New York: Foster and Stewart Publishing Corp., 1946), x + 282 pp.

Dewey, John. Experience and Education (New York: Macmillan Co., 1946), xii + 116 pp.

Hilgard, Ernest R. Theories of Learning (New York: Appleton Century Crofts, Inc., 1948), x + 409 pp.

Jersild, Arthur T., and others. "Studies of Elementary School Classes in Action, II: Pupil Participation and Aspects of Pupil-Teacher Relationships," Journal of Experimental Education, X (December 1941), pp. 119-137.

Johnson, Marguerite Wilker. "Verbal Influences on Children's Behavior," School of Education Monographs in Education, No. 1 (Ann Arbor, Michigan: University of Michigan, 1939), ix + 191 pp.

Lecky, Prescott. Self-Consistency: A Theory of Personality (New York: Island Press, 1945), 154 pp.

Lewin, Kurt. "Psychology and the Process of Group Living," Journal of Social Psychology, XVII (1943), 113-131.

Lewin, Kurt, and others. Changing Behavior and Attitudes. Publication No. 3 of Research Center for Group Dynamics (Cambridge, Massachusetts: Massachusetts Institute of Technology, 1945).

Lippitt, Ronald. "An Analysis of Group Reaction to Three Types of Experimentally Created Social Climate." Unpublished Ph.D. dissertation, University of Iowa, 1940.

Olson, Willard C., and others. "Teacher Personality as Revealed by the Amount and Kind of Verbal Direction Used in Behavior Control," Educational Administration and Supervision, XXIV (February 1938), 81-93.

Snyder, William U. "An Investigation of the Nature of Non-Directive Therapy," Journal of General Psychology, XXXIII (1945), 193-223.

The Psychology of Learning. Edited by Nelson B. Henry. Forty-First Yearbook of the National Society for the Study of Education, Part II (Bloomington, Illinois: Public School Publishing Co., 1942), xiv + 463 pp.

Wickman, E. K. Children's Behavior and Teachers' Attitudes (New York: Commonwealth Fund, 1929), x + 868 pp.

# AN EXPERIMENTAL STUDY OF THE EFFECTIVENESS OF FILMS AND FILMSTRIPS IN TEACHING GEOMETRY*

DONOVAN A. JOHNSON

University of Minnesota

## The Problem

The general objective of the experiment was to investigate the claim that instruction in geometry which emphasizes visualization and application by the use of sound motion pictures or filmstrips is more efficacious in promoting learning than instruction which uses other supplementary aids. Specifically, this study evaluated experimentally the relative effectiveness of certain audio-visual aids as contrasted with conventional instruction on (1) informational learning, (2) development of fundamental skills, (3) ability to make applications of geometric principles and facts, and (4) retention of these types of learning. Wherever the audio-visual aids were used in experimental classes, they replaced an equivalent amount, in time, of conventional instruction.

The null hypothesis is: There is no difference between experimental and control classes in informational learning, problem solving skills, applicational learning, and retention of learning resulting from the use of certain audio-visual aids in classroom instruction.

## The Design of the Experiment

The three requirements for a scientifically designed experiment from which clear cut conclusions may be drawn are: (1) randomization, (2) replication, and (3) local control.

In the present investigation, the selection of the sample as well as the selection of the treatment was made in a random manner. By the use of Tippett's random numbers, twelve schools of all the schools within a radius of 75 miles of Minneapolis and having at least two sections of plane geometry were selected to participate in the experiment. The school enrollments of the participating schools range from 230 to 1500 and the enrollment in participating geometry classes ranged from 14 to 35. Similarly the selection of the type of visual aid to be used was made in a random manner. It is randomization that makes possible an unbiased estimate of the effect and gives validity to statistical tests of significance.

The requirement of replication is satisfied in this experiment in that the investigation consisted of fifteen separate and independent experiments with a total of twenty-seven classes participating. Replication provides the only means of providing the experiment with its own appropriate estimate of error and of diminishing experimental error since the error is inversely proportional to the square root of N where N equals the number of replicates. The number of cooperating schools in this experiment furnishes a broad base for drawing conclusions and guards against accepting generalizations uncritically which often results from the single isolated experiment. It will be noted in the results of the statistical analysis that the outcomes are not the same in all schools. The selection of a few specific schools for this experiment would have yielded results that would have been unwarranted by the findings of the comprehensive investigation. If experiments are to contribute to educational knowledge, significant educational problems should be attacked on a broad base.

Each of the fifteen experimental units was a self-contained experiment since each had its own experimental group and control group. The evidence was determined for each experiment and collated in certain cases. These local control groups eliminate the errors resulting from the use of experimental and control groups each in different localities or under different teachers. Registration methods in each of the participating schools provided for random, heterogeneous grouping of pupils and thus prevented bias from entering into the composition of any of the classes. By a random selection one of the classes in each school was designated as the control class and others as experimental classes. The experimental classes in each school specify the groups which used certain visual aids as a part of their regular instruction. The control class specifies the group taught by the same teacher in the same school but for which no films or filmstrips were provided. The con-

* A summary of a thesis submitted to the faculty of the Graduate School of the University of Minnesota in partial fulfillment of the requirements for the degree Doctor of Philosophy, 1948.

# AN EXPERIMENTAL STUDY OF THE EFFECTIVENESS OF FILMS AND FILMSTRIPS IN TEACHING GEOMETRY*

DONOVAN A. JOHNSON

University of Minnesota

## The Problem

The general objective of the experiment was to investigate the claim that instruction in geometry which emphasizes visualization and application by the use of sound motion pictures or filmstrips is more efficacious in promoting learning than instruction which uses other supplementary aids. Specifically, this study evaluated experimentally the relative effectiveness of certain audio-visual aids as contrasted with conventional instruction on (1) informational learning, (2) development of fundamental skills, (3) ability to make applications of geometric principles and facts, and (4) retention of these types of learning. Wherever the audio-visual aids were used in experimental classes, they replaced an equivalent amount, in time, of conventional instruction.

The null hypothesis is: There is no difference between experimental and control classes in informational learning, problem solving skills, applicational learning, and retention of learning resulting from the use of certain audio-visual aids in classroom instruction.

## The Design of the Experiment

The three requirements for a scientifically designed experiment from which clear cut conclusions may be drawn are: (1) randomization, (2) replication, and (3) local control.

In the present investigation, the selection of the sample as well as the selection of the treatment was made in a random manner. By the use of Tippett's random numbers, twelve schools of all the schools within a radius of 75 miles of Minneapolis and having at least two sections of plane geometry were selected to participate in the experiment. The school enrollments of the participating schools range from 230 to 1500 and the enrollment in participating geometry classes ranged from 14 to 35. Similarly the selection of the type of visual aid to be used was made in a random manner. It is randomization that makes possible an unbiased estimate of the effect and gives validity to statistical tests of significance.

The requirement of replication is satisfied in this experiment in that the investigation consisted of fifteen separate and independent experiments with a total of twenty-seven classes participating. Replication provides the only means of providing the experiment with its own appropriate estimate of error and of diminishing experimental error since the error is inversely proportional to the square root of N where N equals the number of replicates. The number of cooperating schools in this experiment furnishes a broad base for drawing conclusions and guards against accepting generalizations uncritically which often results from the single isolated experiment. It will be noted in the results of the statistical analysis that the outcomes are not the same in all schools. The selection of a few specific schools for this experiment would have yielded results that would have been unwarranted by the findings of the comprehensive investigation. If experiments are to contribute to educational knowledge, significant educational problems should be attacked on a broad base.

Each of the fifteen experimental units was a self-contained experiment since each had its own experimental group and control group. The evidence was determined for each experiment and collated in certain cases. These local control groups eliminate the errors resulting from the use of experimental and control groups each in different localities or under different teachers. Registration methods in each of the participating schools provided for random, heterogeneous grouping of pupils and thus prevented bias from entering into the composition of any of the classes. By a random selection one of the classes in each school was designated as the control class and others as experimental classes. The experimental classes in each school specify the groups which used certain visual aids as a part of their regular instruction. The control class specifies the group taught by the same teacher in the same school but for which no films or filmstrips were provided. The con-

* A summary of a thesis submitted to the faculty of the Graduate School of the University of Minnesota in partial fulfillment of the requirements for the degree Doctor of Philosophy, 1948.

trol classes were given conventional instruction for an equivalent period of time which the experimental classes spent viewing films or filmstrips. Except for the visual materials used in the experimental classes, the experimental and control classes in each school had the same class time, activities, materials and tests. Each teacher was given extensive instruction, orally and in writing, as well as supervision in order to assure uniformity in the use of the visual materials in the classes. A log sheet for each lesson was completed by each teacher so that differences in instruction could be taken into consideration in interpretations.

The instructional units used in the plane geometry experiment were chosen to take advantage of the potentialities of visual aids and at the same time be representative of mathematics subject matter. The instructional units in plane geometry were Circles and Loci. With the circle so closely related to the wheel and locus defined as the path of a point moving according to given conditions, both of these units contain material in which motion is a factor.

In preparing the cooperating teachers for the instruction of these units, a personal conference was held with each, the procedure to follow was outlined in written form, and the teacher was visited during the course of the experiment. In most schools this personal visit included visiting the geometry classes.

The procedure followed in the teaching of each experimental unit was outlined as follows: (1) Prior to the beginning of an experimental unit, a pretest was administered to the participating classes; (2) films or filmstrips were used in the experimental classes at the time and in the manner prescribed in the visual aids teaching guide; (3) control classes participated in worthwhile learning activities as a substitute for the visual aids being used by the experimental classes; (4) experimental and control classes spent the same amount of time learning the experimental units; (5) excluding the films and filmstrips, the control and experimental classes had the same instruction, the same assignments, the same supplementary materials or projects, the same activities and the same tests; (6) each participating instructor was free to use any method, materials, assignments, activities, amount of time, or tests she desired for the experimental classes; (7) at the end of the unit, all classes were given an objective achievement test; (8) approximately two months after the completion of an experimental unit, the achievement test was repeated as a retention test. It is to be noted that the the units were not outlined and prescribed with respect to such items as objectives, subject matter content, time schedule, materials of instruction, or student activities.

Otis Mental Ability Tests were administered to all students who participated in the study and the results of these tests were used as one of the independent matching variables. Achievement in the two instructional units in plane geometry was measured by achievement tests which were constructed by the writer. These achievement tests were used at the beginning of the experiments as pre-tests and again at the end of the experiments as post-tests. The achievement tests were also used as retention tests in some of the experiments. The tests were mailed to reach the school just prior to the time of administration and were returned to the writer by express as soon as possible after they had been administered. The tests were scored and results sent to the teachers within a few days.

The evaluation was concerned with two types of visual aids, namely, sound films and filmstrips. The films[1] used were Circles, Angles and Arcs of Circles, Chords and Tangents of Circles, and Locus. The filmstrips[2] used were Postulates-Triangles and Circles, Vocabulary-Circle I, Vocabulary-Circle II, and Locus. Learning guides were constructed so that the method of using films and filmstrips could be held as uniform as possible. These guides prescribed the time and method of using the teaching aids including discussion questions and a short test. This was done to ensure the use of the films in the manner found most effective by previous research.

In order to compare the effectiveness of films and filmstrips, the plane geometry classes were randomly assigned visual aids as follows: (1) Five experimental classes used only filmstrips, (2) seven experimental classes used only sound films, and (3) three experimental classes used filmstrips and sound films. Three schools each had three geometry classes and, thus, used filmstrips in one class, sound films in another, while the third class was the control.

### The Statistical Analyses

The students who participated in the experiments were given mental ability tests, achievement pre-tests, and achievement post-tests, and where possible achievement retention tests. The scores from these tests were used as the

1. Knowledge Builders, 625 Madison Avenue, New York City.

2. Curriculum Films, Inc., RKO Building, New York City.

basis for the statistical analyses.

In order to illustrate the specific statistical techniques used in this study, data from two of the experimental units will be presented and analysed. The statistical techniques which are widely known by research workers will not be elaborated. The analysis of variance and covariance and the Johnson-Neyman technique are less familiar and thus will be discussed in some detail. The data selected for analysis will involve the data from two schools in which the experimental classes used three films and three filmstrips as a supplement of the regular instruction. The control classes used no films or filmstrips.

The Johnson-Neyman Technique.--The significance of the difference between means was computed for each independent experiment and is presented in Table I.

The probability of the F value was required to be greater than .01 before applying the following formula:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left(\frac{\Sigma x_1^2 + \Sigma x_2^2}{N_1 + N_2 - 2}\right)\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}$$

It will be noted that both of the t values are significant at the one percent level. This shows that the differences in means are statistically significant and in favor of the experimental groups.

Before pooling the data, the homogeneity of the variances was tested separately for the experimental and control groups. Since there were only two schools involved, the homogeneity of variances was tested by Snedecor's F ratio. The significance of the difference between means of experimental and control classes was tested separately by analysis of variance. The obtained F values were less than the tabled values at the one percent level of significance. Since both the experimental and control groups did not differ in variability and the means of the respective groups did not differ significantly, the data were pooled for each of the groups, respectively, before applying the Johnson-Neyman technique. This technique was used to compare the pooled experimental classes with the pooled control classes on Part I Circle Retention-Test. This technique made it possible, (1) to test the significance of the difference between the adjusted means of the experimental group and the control group as based on results obtained on Part I Circle Retention-Test, and (2) to set up a region of significance if a significant difference was found to exist between the experimental and control groups.

The hypothesis $H_1$ tested in this case was the null hypothesis. It stated that there was not a significant difference between the means of the experimental group and the control group in the retention of the informational learning as measured by Part I Circle Retention-Test when the retention-test scores were adjusted on the basis of the Otis Quick-Scoring Mental Ability Test and the Part I Circle Achievement Pre-Test scores.

The basic values required are as follows for Schools F and K:

X = Mental Test Scores
Y = Pre-test Scores of Part I Circle Achievement Test
Z = Retention-Test Scores of Film and Filmstrip Group on Part I Circle Achievement Test
U = Retention-Test Scores of Control Group

| Experimental Group | Control Group |
|---|---|
| $N_1 = 51$ | $N_2 = 44$ |
| $X_1 = 54.78$ | $X_2 = 52.40$ |
| $Y_1 = 9.45$ | $Y_2 = 9.04$ |
| $Z_1 = 23.45$ | $U = 17.04$ |
| $S_{x_1} = 9.81$ | $S_{x_2} = 8.48$ |
| $S_{y_1} = 4.51$ | $S_{y_2} = 4.49$ |
| $S_z = 6.42$ | $S_u = 7.88$ |
| $r_{x_1 y_1} = .48$ | $r_{x_2 y_2} = .46$ |
| $r_{x_1 z} = .61$ | $r_{x_2 u} = .30$ |
| $r_{y_1 z} = .66$ | $r_{y_2 u} = .45$ |

The ratio between the relative and absolute minimum sums of squares provides a rigorous test of significance of the differences between means. In our case this ratio which is designated as w was found to be 3.934. By interpolation from Table V in Johnson and Neyman's publication[3], we obtain $W_{.01} = 12.850$. Thus,

3. Palmer O. Johnson and J. Neyman, "Tests of Certain Linear Hypotheses and Their Application to Some Educational Problems," Statistical Research Memoirs, I (1936), pp. 57-93.

TABLE I

RETENTION-TEST RESULTS WITH FILMS AND FILMSTRIPS--TESTS OF SIGNIFICANCE OF DIFFERENCES BETWEEN MEANS OF EXPERIMENTAL AND CONTROL CLASSES FOR PART I KNOWLEDGE OF FACTS CIRCLE TEST

| School | Class | N | Mean | S.D. | F | F Values at .05 | F Values at .01 | t | Probability |
|---|---|---|---|---|---|---|---|---|---|
| F | Exp. | 27 | 25.0 | 6.56 | | | | | |
| | Con. | 23 | 19.8 | 6.61 | 1.01 | 1.96 | 2.63 | 2.76 | .01 |
| K | Exp. | 24 | 21.7 | 5.92 | | | | | |
| | Con. | 21 | 14.0 | 8.17 | 1.90 | 2.04 | 2.78 | 3.65 | .01 |
| Total | Exp. | 51 | 23.4 | 6.42 | | | | | |
| | Con. | 44 | 17.0 | 7.88 | 1.50 | 1.63 | 2.00 | | |
| | Total | 95 | 20.4 | 7.79 | | | | | |

TABLE II

TEST OF HOMOGENEITY OF REGRESSION

| | d.f. | Sum of Squares | Variance | F |
|---|---|---|---|---|
| Deviations from Average Regression Within | 91 | 2394.1784 | 26.3097 | |
| Deviations from Individual Regressions | 90 | 2362.5955 | 26.2511 | 1.20 |
| Deviations Among Group Regressions | 1 | 31.5829 | 31.5829 | |

since the observed value of W is less than the interpolated table value of $W_{.01}$, we conclude that the experimental and control groups do differ significantly at the one percent level on Part I Circle Retention-Test. In other words, the audio-visual aids group was on the average significantly superior to the non-audio-visual aids group in informational learning as measured by Part I Circle Retention-Test, where differences in pre-test and mental test scores were controlled.

The region of significance which is represented by an ellipse is shown in Figure 1. To the right of the line of non-significance, the value of $F(x',y')$ is favorable to the experimental groups. The value is significant at the one percent level for that part of the population enclosed within the ellipse. That is for all values of $(x', y')$, the basic characters of matching--mental ability and initial information--within the region of significance, the hypothesis $H_1$ $(x', y')$ is rejected. To the left of the line of non-significance the value of $F(x', y')$ is favorable to the control group but nowhere significantly so.

The Analysis of Variance and Covariance.--The analysis of variance and covariance makes it possible to partial out the effects of an initial lack of equivalence of the classes. It makes it possible to conduct an experiment with intact classes without matched pairs being involved. In this experiment mental test scores and pre-test scores were obtained so that it was possible to hold constant the effects of mental ability and initial information.

The analysis of variance and covariance will be illustrated for the data from these same schools for the retention of applicational learning. Similar to the data presented above, the differences between means in the retention of applicational learning were statistically significant in favor of the experimental classes. Since the requirements for pooling the data were satisfied, the analysis is completed for the data of the experimental and control classes pooled separately.

Before an analysis of variance and covariance can be completed, certain basic assumptions must be satisfied. One of these is the assumption of homogeneity of variance within classes. This homogeneity of variance can be tested by the Welch Nayer $L_1$ test if more than two classes are involved and by the F test if only two classes are being analysed. For the data of this analysis, the $L_1$ value, .958, exceeds the tabled value at the 5 percent level of significance and, thus, the condition of homogeneity of variance was satisfied.

Another basic assumption underlying the analysis of covariance is that the regression coefficients within groups are homogeneous. For the data at hand, the analysis of variance was used to test the homogeneity of the regression coefficients. As shown in Table II, the F ratio of 1.20 was not significant and, thus, the condition of homogeneity of regression within groups was satisfied. This assumption may also be tested by the Welch-Nayer $L_1$ test.

The analysis of variance and covariance summarized in Table III indicates that the difference between the mean of the pooled experimental group, and the mean of the pooled control groups is significant at the 1 percent level when the factors of mental ability and initial information are held constant. The difference is in favor of the experimental or audio-visual groups. Hence we can reject the hypothesis that there is no difference between the means of the experimental and control groups in the retention of applicational learning resulting from the use of the films and filmstrips as a complement of conventional instruction.

## Summary

Since the scope of this experiment precludes the possibility of presenting all the data and analyses, the results are summarized in Tables IV and V. It will be noted in these tables that very few differences between the means of experimental and control groups were statistically significant. There were differences in favor of experimental groups as well as control groups. These results point out how readily an incorrect conclusion might be drawn if a single experiment were conducted. Thus, it would seem important that cooperative experimentation is essential to furnish a broad base for conclusions and guard against accepting generalizations uncritically.

The one area in which the results were consistently in favor of the experimental groups was in the retention of learning in those classes using three films and three filmstrips in the circle unit. These results have added significance when it is noted that these experimental classes were not significantly different on the post-test means.

## Implications

In conclusion we wish to summarize by indicating certain implications which appear warranted as a result of this study. It should be noted that the study was, (1) designed to measure the relative effectiveness of certain audio-visual materials in terms of unit objectives; (2) conducted under as nearly normal class-

TABLE III

ANALYSIS OF VARIANCE AND COVARIANCE OF PART IV CIRCLE RETENTION-TEST SCORES FOR SCHOOLS A AND K WITH MENTAL TEST SCORES AND PRE-TEST SCORES HELD CONSTANT

A. Data for Analysis of Variance and Covariance

x=Mental Test y=Pre-Test z=Retention-Test

| | N | $x^2$ | xy | $y^2$ | yz | $z^2$ | xz |
|---|---|---|---|---|---|---|---|
| Exp. | 51 | 4820.6275 | 694.0980 | 1537.9216 | 611.4706 | 1857.1765 | 1194.4118 |
| Con. | 44 | 3096.6363 | 299.5909 | 602.9773 | 85.7955 | 1039.1591 | 561.3182 |
| | | (7917.2638) | (993.6889) | (2140.8989) | (697.2661) | (2896.3356) | (1755.7300) |
| Total | 95 | 8050.5263 | 1069.4211 | 2183.9368 | 842.6632 | 3387.5368 | 2011.5789 |

B. Adjustment Table for Analysis of Covariance

| | Denominator | $b_1$ | $b_2$ | $b_1$ xz | $b_2$ yz | $z^2$ | d.f. |
|---|---|---|---|---|---|---|---|
| Exp. | 6931975.124 | .203765 | .305632 | 243.37932 | 186.88498 | 1426.91220 | 48 |
| Con. | 1777446.688 | .175959 | .054861 | 98.76899 | 4.70683 | 935.68328 | 41 |
| | | | | | | (2362.59548) | |
| | 15962643.73 | .192072 | .236539 | 337.22657 | 164.93063 | 2394.17840 | |
| Total | 16438179.16 | .212432 | .281823 | 427.32373 | 237.48187 | 2722.73120 | 89 |

C. Adjusted or Reduced

| Source of Variation | d.f. | Sum of Squares | Mean Square | F |
|---|---|---|---|---|
| Between Groups | 1 | 328.5528 | 328.5528 | |
| Within Groups | 91 | 2394.1784 | 26.3097 | 12.48 |
| Total | 92 | 2722.7312 | | |

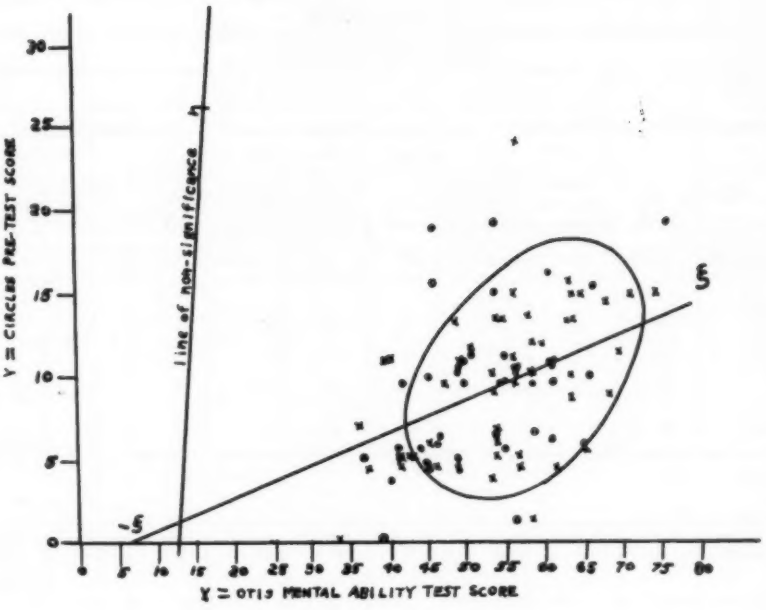


**Figure 1**

**Comparison of the experimental students (x) and the control students (o) in achievement on Part I, Knowledge of Facts, Circles Retention-Test, when matching was affected on the Otis Quick-Scoring Mental Ability Test and Part I Circles Pre-Test.**

TABLE IV

SUMMARY OF THE DIFFERENCES BETWEEN EXPERIMENTAL AND CONTROL CLASSES ON POST-TESTS

| Circle Unit 3 Films 3 Filmstrips | Experimental treatment | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Film and Filmstrip | | | Films | | | Filmstrips | | |
| Type of Learning | No Sig. Differ. | Exper. Groups Superior | Control Groups Superior | No Sig. Differ. | Exper. Groups Superior | Control Groups Super. | No Sig. Differ. | Exper. Groups Superior | Control Groups Superior |
| Informational | 3 | 0 | 0 | 6 | 1 | 0 | 5 | 0 | 0 |
| Problem Solving | 3 | 0 | 0 | 7 | 0 | 0 | 5 | 0 | 0 |
| Applica-tional | 2 | 1 | 0 | 5 | 2 | 0 | 4 | 0 | 1 |
| Total | 8 | 1 | 0 | 18 | 3 | 0 | 14 | 0 | 1 |
| Locus Unit 1 Film 1 Filmstrip | | | | | | | | | |
| Informational | 3 | 0 | 0 | 7 | 0 | 0 | 5 | 0 | 0 |
| Problem Solving | 2 | 1 | 0 | 6 | 0 | 1 | 5 | 0 | 0 |
| Applica-tional | 3 | 0 | 0 | 6 | 0 | 1 | 5 | 0 | 0 |
| Total | 8 | 1 | 0 | 10 | 0 | 2 | 15 | 0 | 0 |

TABLE V

SUMMARY OF THE DIFFERENCES BETWEEN EXPERIMENTAL AND CONTROL CLASSES ON RETENTION-TESTS

| Circle Unit 3 Films 3 Filmstrips | Experimental treatment | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Film and Filmstrip | | | Films | | | Filmstrips | | |
| Type of Learning | No Sig. Differ. | Exper. Group Superior | Con. Group Superior | No Sig. Differ. | Exper. Group Super. | Con. Group Super. | No Sig. Differ. | Exper. Group Super. | Con. Group Super. |
| Informational | 0 | 2 | 0 | 3 | 0 | 0 | 1 | 1 | 0 |
| Problem Solving | 0 | 2 | 0 | 3 | 0 | 0 | 2 | 0 | 0 |
| Applica-tional | 0 | 2 | 0 | 3 | 0 | 0 | 1 | 0 | 1 |
| Total | 0 | 6 | 0 | 0 | 0 | 0 | 4 | 1 | 1 |
| Locus Unit 1 Film 1 Filmstrip | | | | | | | | | |
| Informational | 2 | 0 | 0 | 2 | 0 | 1 | 2 | 0 | 0 |
| Problem Solving | 2 | 0 | 0 | 3 | 0 | 0 | 2 | 0 | 0 |
| Applica-tional | 2 | 0 | 0 | 2 | 0 | 1 | 2 | 0 | 0 |
| Total | 6 | 0 | 0 | 7 | 0 | 2 | 6 | 0 | 0 |

room situations as possible; (3) conducted at two high school levels; (4) made with three types of audio-visual aids; and (5) made in schools which were selected in a random manner. It should be further noted that this study attempted to evaluate the relative effectiveness of audio-visual material as they affected informational learning, the learning of problem solving skills, applicational learning and the retention of learning in these areas. The implications are as follows:

1. It appears that audio-visual aids which are developed for use in mathematics classes might be more effective as aids to learning if they were designed to supplement rather than repeat the type of instruction which the students have in the typical mathematics classroom. The audio-visual aids used in this study contained material which was frequently similar to material presented in mathematics textbooks. Some of the material presented by the films used animated drawings that were similar to the type of illustrations which mathematics teachers customarily draw on the blackboard. Experimental results might be considerably more favorable to the effectiveness of visual materials if the audio-visual aids would bring to the classroom unique experiences that are vital and functional to the adolescent student, that is, different from experiences supplied by textbook or teacher or by other non-visual aids. For example, a motion picture film could show the geometry used in aviation or a filmstrip could illustrate the geometric designs used by nature in snowflakes, bee cells, flowers, seeds or crystals.

2. An implication from this study would indicate that if we are to secure audio-visual aids that actually enhance learning, careful experimental studies should be made testing the claims made for them before they are made available for instructional purposes. The number of experiments which have resulted in significant differences in means have been relatively few. And even when statistically significant results were obtained, one might raise the question as to whether the results are pedagogically significant. Additional experimentation is also needed to test the effectiveness of other types of materials of instruction, such as charts, opaque projections, and stereographs.

3. On the basis of the cooperation obtained from the administrators and teachers of participating schools, it appears that experimentation in randomly selected schools is entirely feasible. The teachers who participated exhibited considerable interest in the experiment and cooperated in following directions and suggestions in a highly satisfactory manner. However, most of the participating teachers lacked training and experience in methods of research and in methods of using visual materials. Furthermore, the time available to teachers in view of the usually heavy teaching load limits the extent to which they can prepare adequate lessons using available materials. There is also a need for the coordination of means of informing teachers of the materials that are available. Many of the teachers participating in this study had never used or seen a mathematics film or filmstrip before.

# AN EXPERIMENTAL EVALUATION OF THE RELATIVE EFFECTIVENESS OF CERTAIN AUDIO-VISUAL AIDS IN VOCATIONAL AGRICULTURE*

RALPH R. BENTLEY

Purdue University

The purpose of this experimental study was, (1) to determine the relative effectiveness of certain audio-visual aids on informational learning and on the ability to solve problems or make application of principles in vocational agriculture, and (2) to determine the relative effectiveness of certain audio-visual aids on the retention of these two types of learning in vocational agriculture.

## The Design of the Experiment

The selection of the cooperating schools.--The schools which cooperated in this study were selected with the aid of the Agricultural Education staff of the University of Minnesota and the Minnesota State Department of Agricultural Education. The cooperating schools were located in the southern half of Minnesota and were selected on the basis of the following criteria: (1) the agricultural area in which the school was located; (2) the ability and success of the local teacher of agriculture and his interest and willingness to cooperate in the proposed experimental study; (3) the number of students enrolled in the agriculture classes of each school; and (4) the cooperation which might be expected from the administrators in each school.

The selection of the experimental projects.--The instructional units which served as the basis for the three experimental projects were in home gardening, swine production, and pasture production. The units were selected because (1) they were generally taught to classes in vocational agriculture; (2) suitable audio-visual aids were available for these units; (3) it was possible to measure the effectiveness of audio-visual aids in both crop and livestock enterprises; and (4) these units were usually taught to students of different ages.

The selection of the audio-visual aids.--The audio-visual aids used in this study were selected by the writer who was assisted by audio-visual education specialists of Iowa State College and of the University of Minnesota. The following factors were taken into consideration in selecting the audio-visual aids: (1) the educational level for which they were suited; (2) the subject-matter content; (3) the photography and technical construction, (4) the color of the pictures; and (5) the physical condition of the films and slides.

The independent experiments.-- The investigation was made in eight different schools and consisted of twelve independent experiments. Each of the twelve independent experiments had its own experimental and control groups. Six of the experiments were in the Home Garden Project, four in the Swine Production Project and two in the Pasture Production Project. Although the independent experiments, in each of the three experimental projects were complete experiments within themselves there were certain features which were common to the independent experiments in each experimental project. The common features were, (1) all students in a given experimental project were taught the same instructional unit; (2) all students in a given experimental project were administered the same tests with the exception of the retention test which was not administered in certain independent experiments; (3) the experimental groups in each of the experimental projects had the same audio-visual aids as part of their instruction; and (4) the instructional time, in each of the independent experiments, was equal for the experimental and control groups.

The experimental and control groups.--In each independent experiment there was an experimental or audio-visual aids group and a control or non-audio-visual aids group. Where two sections of the same agriculture class were taught in the same school, one section was used as the experimental group and the other as the control group. The treatment in each case was determined by lot. Most of the experiments, however, were conducted in schools where there was only one class section. In such instances the experimental and control groups were obtained by dividing the classes at random. The instruction of both groups, in each of the

* A summary of a thesis submitted to the faculty of the Graduate School of the University of Minnesota in partial fulfillment of the requirements for the degress Doctor of Philosophy, 1947.

independent experiments, where divided classes were used, was under the direct supervision of the same instructor and the groups were taught as a unit except when the audio-visual aids were shown.

The testing program.--The Terman-McNemar Test of Mental Ability was administered to all students who participated in this experimental study and the results were used as one of the independent matching variables. Achievement in the three experimental projects, Home Gardening, Swine Production, and Pasture Production, was measured by achievement tests which were constructed by the writer and were designed to measure in terms of the objectives of each respective unit. These tests were used before the instructional period as pre-tests, at the end of the instructional period as post-tests in all experimental projects and as retention tests in certain of the Home Garden and Swine Production experiments.

## The Statistical Analysis

The statistical techniques used in this study are as follows: (1) the arithmetic mean was used as a measure of central tendency and the standard deviation and variance were used as measures of variability. (2) When it was necessary to test the homogeneity of data for two groups, the t-test was used to determine the significance of the difference between means, and the F-test was used to determine the significance of the difference between variances. (3) When it was necessary to test the homogeneity of data for more than two groups, the analysis of variance was used to determine the significance between means and the Welch-Nayer ($L_1$ test) was used to determine the significance of the difference between variances. (4) The data for the independent experimental groups in each experimental project were pooled whenever it was shown that their means did not differ significantly and that their variances were homogeneous. The control groups were treated similarly. (5) The analysis of variance and covariance was used to test the significance of the difference between the means of the experimental and control groups on the retention test. (6) The Johnson-Neyman technique[1] was used to compare the pooled experimental groups and the pooled control groups in the Home Garden and the Swine Production experimental projects. This technique makes possible the comparison of the means of the two groups after their adjustment for any differences in the basic characters of matching which in this experimental study were the Terman-McNemar Test of Mental Ability scores and the achievement pre-test scores. If there proves to be a non-significant difference between the means, the analysis is not carried further. However, if there is a significant difference between the means after adjustment then the analysis is continued to set up a region of significance. The region of significance enables one to determine all the systems of matching variables for which a difference would be found between the experimental and control groups. That is, it determines for what kind of students as defined by the matching variables a significant difference would be found in favor of one or the other of the experimental and control groups.

## The Home Garden Experimental Groups

The Home Garden Experimental Project consisted of six independent experiments each of which had its own experimental and control groups that were taught by the same instructor, used the same reference materials, spent an equal amount of time on the home garden unit and were administered the same tests at the same time. The instruction of the two groups in each experiment was identical except that the control group studied and discussed the home garden unit for a time equal to that which was used in presenting the movie to the experimental group. The movie entitled "The Farm Garden" and the film strips "Gardening for Victory" Parts I and II were the audio-visual aids used in the Home Garden Project. Achievement was measured in this experimental project by the Home Garden Achievement Test of which Part I was designed to measure informational learning and Part II applicational learning. On the basis of the separate post-test results for Part I and Part II the Data of each part were pooled for the six independent experimental groups when it was shown that their variances were homogeneous and that the differences among their means were not significant. This was likewise true of the variances and means of the six independent control groups.

The Application of the Johnson-Neyman Technique.--The Johnson-Neyman technique was used on the pooled data (1) to test the significance of the difference between the adjusted means of the audio-visual aids or experimental group and the non-audio-visual aids or control group as based on results obtained on Part I of the Home Garden Achievement Post-test, and (2) to set up a region of significance if a significant difference was found to exist between the experimental and control groups. Table I contains the basic statistics which were used in making application of the Johnson-Neyman technique to the results of Part I of the Home Garden Achiev-

1. Palmer O. Johnson and J. Neyman, "Tests of Certain Linear Hypotheses and Their Application to Some Educational Problems," Statistical Research Memoirs, I (1936), pp. 57-93.

TABLE I

STATISTICS COMPUTED FOR THE APPLICATION OF THE JOHNSON-NEYMAN METHOD OF ANALYSIS OF HYPOTHESIS $H_1$

Calculated Data

1. Notations
   - Dependent Variable — $Z_s$: Part I Home Garden Achievement Post-test
   - Independent Variables — $X_s$: Part I Home Garden Achievement Pre-Test; $Y_s$: Terman-McNemar Test of Mental Ability
   - Frequencies — $N_s$

   where s = 1, 2 denote Experimental and Control Data, respectively

2. Results

| | |
|---|---|
| $N_1 = 73$ | $N_2 = 73$ |
| $\Sigma Z_1 = 2145$ | $\Sigma Z_2 = 2038$ |
| $\Sigma Z_1^2 = 65235$ | $\Sigma Z_2^2 = 59592$ |
| $\Sigma X_1 = 1739$ | $\Sigma X_2 = 1710$ |
| $\Sigma X_1^2 = 44001$ | $\Sigma X_2^2 = 42584$ |
| $\Sigma Y_1 = 6527$ | $\Sigma Y_2 = 6599$ |
| $\Sigma Y_1^2 = 619977$ | $\Sigma Y_2^2 = 637364$ |
| $\Sigma Z_1 X_1 = 52415$ | $\Sigma Z_2 X_2 = 49346$ |
| $\Sigma Z_1 Y_1 = 197076$ | $\Sigma Z_2 Y_2 = 189350$ |
| $\Sigma X_1 Y_1 = 161021$ | $\Sigma X_2 Y_2 = 159117$ |
| Mean $\bar{Z}_1 = 29.38$ | Mean $\bar{Z}_2 = 27.92$ |
| S. D. $Z_1 = 5.54$ | S. D. $Z_2 = 6.12$ |
| Mean $\bar{X}_1 = 23.82$ | Mean $\bar{X}_2 = 23.42$ |
| S. D. $X_1 = 5.84$ | S. D. $X_2 = 5.93$ |
| Mean $\bar{Y}_1 = 89.41$ | Mean $\bar{Y}_2 = 90.40$ |
| S. D. $Y_1 = 12.02$ | S. D. $Y_2 = 12.95$ |

ement Post-test.

The hypothesis $H_1$ tested in this case was the null hypothesis. It states that there is not a significant difference between the means of the experimental or audio-visual aids group and the control or non-audio-visual aids group in informational learning as measured by Part I of the Home Garden Achievement Post-test, when the post-test scores are adjusted on the basis of the Terman-McNemar Test of Mental Ability and Part I of the Home Garden Achievement Pre-test scores.

The ratio between the relative and absolute minimum sums of squares provides a rigorous test of significance of the differences between means. In our case, this ratio which is designated as u was found to be .8134. By interpolation from Table V in Johnson and Neyman's article[2] we obtain $u_{.01} = .9112$. Thus, since our observed value of u is less than the interpolated value of $u_{.01}$, we conclude that the experimental and control groups do differ significantly at the 1 percent level on Part I of the Home Garden Achievement Post-test. In other words, the audio-visual aids group was on the average significantly superior to the non-audio-visual aids group in informational learning as measured by Part I of the Home Garden Achievement Post-test.

The region of significance which is represented by a hyperbola is shown in Figure 1. Since there are two branches of the hyperbola we have the interesting situation whereby for every system of values $X'Y'$ in the left branch of the hyperbola the $H_1(X'Y')$ is rejected in favor of the experimental group, and for every system of the values $X'Y'$ in the right branch of the hyperbola the $H_1(X'Y')$ will be rejected in favor of the control group. Study of the region of significance indicates that the experimental group may be said to be superior to the control group throughout the range of mental ability of the students and for the average or below average student as measured by Part I of the pre-test.

The Johnson-Neyman technique was also used to test the significance of the difference between the adjusted means of the experimental and control groups as based on the results obtained from Part II of the Home Garden Achievement Post-test. It will be recalled that Part II was designed to test applicational learning. Table II gives the basic statistics which were used in applying the Johnson-Neyman technique to the results of Part II of the Home Garden Achievement Post-test.

The ratio between the relative and absolute minimum sums of squares which has been designated as u, is .9683. By interpolation from Table V in Johnson-Neyman's article[3] we obtain $u_{.01} = .9112$. Since our observed value of u is greater than the interpolated table value for $u_{.01}$ we conclude that the adjusted means of the experimental and control groups do not differ significantly at the 1 percent level of significance. Thus, the audio-visual aids group and the non-audio-visual aids group were equal in applicational learning or in their ability to solve problems as measured by Part II of the Home Garden Achievement Post-test.

The Home Garden Achievement Retention-test was administered in two of the independent experiments, and the results were analyzed by the analysis of variance and covariance with the post-test results held constant. The difference between the means of the experimental and control groups for Part I of the Home Garden Achievement Retention-test was not significant. This was likewise true of the experimental and control groups as measured by Part II of the Home Garden Achievement Retention-test.

## The Swine Production Experimental Project

The Swine Production Experimental Project consisted of four independent experiments, each having its own experimental or audio-visual aids group and control or non-audio-visual aids group. The groups in each independent experiment were taught by the same instructor, used the same reference materials, spent the same length of time on the swine unit, and were administered the same tests at the same time.

In each of the independent experiments the experimental and control groups were taught as a unit except for time necessary to show the movie "Pig Projects Make Profits", to the experimental group. During the time the movie was being shown the control group continued its study of the swine unit as directed by the local teacher of agriculture.

The Swine Production Achievement Test was administered as pre-test and post-test in all independent experiments and as retention test in three of the independent experiments.

It was statistically permissible to pool the data of the four independent experimental groups. This was also true of the four control groups. The Johnson-Neyman technique was used to test the significance of the difference between the post-test means of the pooled experimental

2. Johnson-Neyman, op. cit.

3. Johnson-Neyman, op. cit.

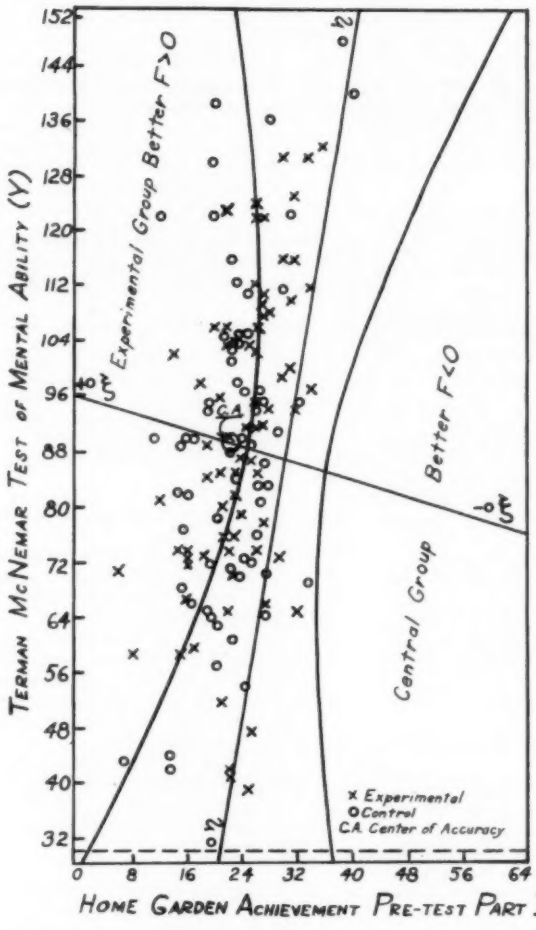


FIGURE 1.—Region of Significance Represented by a Hyperbola. Comparison of the experimental students (x) and the control students (o) in achievement on Part I of the Home Garden Achievement Post-test when matching was effected on the Terman-McNemar Test of Mental Ability and Part I of the Home Garden Achievement Pre-test.

TABLE II

STATISTICS COMPUTED FOR THE APPLICATION OF THE JOHNSON-NEYMAN METHOD OF ANALYSIS OF HYPOTHESIS $H_{II}$

Calculated Data

1. Notations
   - Dependent Variable: $Z_s$: Part II Home Garden Achievement Post-test.
   - Independent Variables: $X_s$: Part II Home Garden Achievement Pre-test.; $Y_s$: Terman-McNemar Test of Mental Ability
   - Frequencies: $N_s$

   where s = 1, 2 denote Experimental and Control Data, respectively.

2. Results

| | |
|---|---|
| $N_1 = 73$ | $N_2 = 73$ |
| $\Sigma Z_1 = 1960$ | $\Sigma Z_2 = 1826$ |
| $\Sigma Z_1^2 = 63592$ | $\Sigma Z_2^2 = 52260$ |
| $\Sigma X_1 = 1695$ | $\Sigma X_2 = 1601$ |
| $\Sigma X_1^2 = 47679$ | $\Sigma X_2^2 = 40149$ |
| $\Sigma Y_1 = 6527$ | $\Sigma Y_2 = 6599$ |
| $\Sigma Y_1^2 = 619977$ | $\Sigma Y_2^2 = 637364$ |
| $\Sigma Z_1X_1 = 50024$ | $\Sigma Z_2X_2 = 42734$ |
| $\Sigma Z_1Y_1 = 183757$ | $\Sigma Z_2Y_2 = 171674$ |
| $\Sigma X_1Y_1 = 155555$ | $\Sigma X_2Y_2 = 145854$ |
| Mean $\bar{Z}_1 = 26.85$ | Mean $\bar{Z}_2 = 25.01$ |
| S. D. $Z_1 = 12.34$ | S. D. $Z_2 = 9.56$ |
| Mean $\bar{X}_1 = 23.22$ | Mean $\bar{X}_2 = 21.93$ |
| S. D. $X_1 = 10.75$ | S. D. $X_2 = 8.36$ |
| Mean $\bar{Y}_1 = 89.41$ | Mean $\bar{Y}_2 = 90.40$ |
| S. D. $Y_1 = 12.02$ | S. D. $Y_2 = 12.95$ |

groups and the pooled control groups. Table III gives the basic statistics which were used in applying the Johnson-Neyman technique to the results of the Swine Production Achievement Post-test.

The ratio between the relative and absolute minimum sums of squares in this case, or u, is .9954. By interpolation from Table V in Johnson-Neyman's article[4] we obtain $u_{.01}$ = .7936. Thus, since our observed value of $u_{.01}$ is greater than the interpolated value we conclude that there is not a significant difference between the means of the experimental group and the control group. In other words the mean achievement of the audio-visual and the non-audio-visual aids groups as measured by the Swine Production Achievement Test was not significantly different.

The results of the Swine Production Achievement Retention-test were analyzed by the analysis of variance and covariance with the post-test results held constant. The difference between the means of the experimental and control groups was not significant as measured by the retention test.

## The Pasture Production Experimental Project

The Pasture Production Experimental Project consisted of two independent experiments in which three sets of kodochrome pasture slides with descriptive notes were used as part of the instruction of the experimental group. The experimental and control groups in each independent experiment were taught as a unit except during the time when the audio-visual aids were shown. The control groups in each experiment were taught separately from the experimental group for a period of time equivalent to that used in showing the slides to the experimental groups. The instruction of both groups in each independent experiment was supervised by the same instructor.

It was statistically permissible to pool the data for the two independent experimental groups and it was also possible to pool the data for the two control groups. Applying the t-test of significance to the pooled data for Part I of the Pasture Production Achievement Post-test results, it was found that the means of the experimental and control groups were not significantly different. This was likewise true of the results for Part II of the Pasture Production Achievement Post-test.

## Conclusions

The conclusions resulting from the three experimental projects in this investigation are as follows:

1. The Home Garden Experimental Project.
   a. The mean of the experimental or audio-visual aids group was significantly higher than the mean of the control or non-audio-visual aids group in informational learning as measured by Part I of the Home Garden Achievement Post-test.
   b. The region of significance indicated that the mean of the experimental group may be said to be significantly superior to the control group throughout the range of mental ability of the students and for the average or below average student as measured by Part I of the Home Garden Achievement Pre-test.
   c. The mean of the experimental group was not significantly different from the mean of the control group in applicational learning as measured by Part II of the Home Garden Achievement Post-test.
   d. In the two independent experiments where retention was measured, the means of the experimental and control groups were not significantly different in either informational or applicational learning as measured by Part I and II of the Home Garden Achievement Retention-test, respectively.
2. The Swine Production Experimental Project.
   a. The mean of the experimental group was not significantly different from the mean of the control group in achievement as measured by the Swine Production Achievement Post-test.
   b. The mean of the experimental group was not significantly different from the mean of the control group in achievement as measured by the Swine Production Achievement Retention-test.
3. The Pasture Production Experimental Project.
   a. The mean of the experimental group was not significantly different from the mean of the control group in in-

4. Johnson-Neyman, op. cit.

TABLE III

STATISTICS COMPUTED FOR THE APPLICATION OF THE JOHNSON-NEYMAN METHOD OF ANALYSIS TO HYPOTHESIS $H_{III}$

Calculated Data

1. Notations

| | |
|---|---|
| Dependent Variable | $Z_s$: Swine Production Achievement Post-test. |
| Independent Variables | $X_s$: Swine Production Achievement Pre-test. |
| | $Y_s$: Terman-McNemar Test of Mental Ability. |
| Frequencies | $N_s$ |

where s = 1, 2 denote Experimental and Control Data, respectively.

2. Results

| | |
|---|---|
| $N_1 = 37$ | $N_2 = 42$ |
| $\Sigma Z_1 = 2117$ | $\Sigma Z_2 = 2337$ |
| $\Sigma Z_1^2 = 126157$ | $\Sigma Z_2^2 = 136455$ |
| $\Sigma X_1 = 1568$ | $\Sigma X_2 = 1819$ |
| $\Sigma X_1^2 = 71224$ | $\Sigma X_2^2 = 85053$ |
| $\Sigma Y_1 = 3295$ | $\Sigma Y_2 = 3936$ |
| $\Sigma Y_1^2 = 310819$ | $\Sigma Y_2^2 = 388898$ |
| $\Sigma Z_1X_1 = 92845$ | $\Sigma Z_2X_2 = 104876$ |
| $\Sigma Z_1Y_1 = 194233$ | $\Sigma Z_2Y_2 = 223957$ |
| $\Sigma X_1Y_1 = 145758$ | $\Sigma X_2Y_2 = 176712$ |
| Mean $\overline{Z}_1 = 57.22$ | Mean $\overline{Z}_2 = 55.67$ |
| S. D. $Z_1 = 11.82$ | S. D. $Z_2 = 12.51$ |
| Mean $\overline{X}_1 = 42.38$ | Mean $\overline{X}_2 = 43.31$ |
| S. D. $X_1 = 11.76$ | S. D. $X_2 = 12.36$ |
| Mean $\overline{Y}_1 = 89.03$ | Mean $\overline{Y}_2 = 93.71$ |
| S. D. $Y_1 = 11.50$ | S. D. $Y_2 = 11.73$ |

formational learning as measured by Part I of the Pasture Production Achievement Post-test.

b. The mean of the experimental group was not significantly different from the mean of the control group in applicational learning as measured by Part II of the Pasture Production Achievement Post-test.

## Implications

In conclusion we wish to summarize by indicating certain implications which appear to be warranted as a result of this study:

1. Audio-visual aids which are developed for use in vocational agriculture classes would likely be more effective as aids to learning if they were designed to supplement rather than repeat the visual experiences the students have had.
2. Audio-visual aids which are designed to teach new ideas or principles should be accompanied with adequate interpretation either by means of sound recording or by the instructor.
3. The catalog descriptions of audio-visual aids were not dependable guides to follow when selecting audio-visual materials which were effective aids to learning in classroom situations.
4. If we are to secure audio-visual aids that actually enhance learning, careful experimental studies should be made, testing their value before they are made available for instructional purposes. An analogy may be made here from the field of agriculture. Namely, no hybrid corn is recommended without a careful experimental measure of its yielding ability. Why shouldn't the same critical attitude be exercised in the selection of audio-visual aids which are to be used in learning situations?

# NOTE ON ESTABLISHING NORMS IN A LARGE CITY SYSTEM

J. A. KEATS

Australian Council for Educational Research

In his article in the March 1948 number of the Journal of Experimental Education, H. L. Stein[1] suggested how the problem of establishing measures of central tendency and variability for a school system could be met. The solution offered for finding a measure of central tendency is of practical interest, although an indication of the extent of the bias caused by heterogeneity of variance in the schools would have been welcome. However, the solution given for an unbiased estimate of population variance is probably not unbiased. The formula quoted from Hoel gives an unbiased estimate of the within schools variance, and since this variance has been shown by Stein to be heterogeneous, i.e. varying by an amount greater than would be expected by chance, there is little point in estimating it. A more exact procedure for obtaining an unbiased estimate of the population variance is given below; but all such procedures must ignore the heterogeneity of school variances.

Since the derivation of the formulae necessary for this procedure is not easy, it has been included in the appendix. However, the ideas behind the mathematics are fairly straightforward, and if they are acceptable the algebra can be verified at leisure. In most surveys of the type quoted it is found that there is a highly significant difference between the means of some of the schools tested. This is not surprising as it is obvious that a school will in the main serve the district in which it is situated and that this district will contain predominantly members within a limited socio-economic range; it is to be expected then that the scores of children of such a school in any given test will often have a mean which will differ from the population mean by more than can be attributed to chance and in general will have a smaller range. Furthermore, the degree of homogeneity will differ from school to school, and we must expect differences in standard deviations which are greater than chance, as has been demonstrated by Stein.

The ideas behind the formulae involve the following steps:

1. To test whether or not the means of schools differ from each other by more than can be expected by chance fluctuation. (Analysis of Variance--test the between schools against the within schools variance.)

2. If a positive result is obtained in 1, above, then to estimate the amount by which the pooled within schools variance has been reduced by the selective factors mentioned above and hence obtain an unbiased estimate of the population variance.

Re-Analysis of the Data Quoted by Stein

Analysis of Variance

| | S. of S. | Ds. of F. | Variance | F |
|---|---|---|---|---|
| Between Schools | 65,620.00 | 45 | 1,458.2222 | 7.2 |
| Within Schools | 426,013.222 | 2095 | 203.3476 ($\sigma$ = 14.26) | |
| Total | 491,013.22 | 2140 | 229.7351 ($\sigma$ = 15.16) | |

Stein has shown that the conditions underlying the F test are not fulfilled, since the schools cannot be regarded as homogeneous with respect to variability. However, an empirical study by Lindquist[2] (Chapter V, Sections 8 and 9) suggests that when the effect of this heterogeneity is taken into account an $F_{45,2095} = 7.2$ would almost certainly be significant at the .1% point.

We now proceed to estimate the amount the variance of the population has been reduced by the selective factors operating.

Let $V_i$ represent the component of variance within schools and $V_{bs}$ the component of variance between schools.

Then $$V_{bs} = \frac{(M_{bs} - M_i)(r - 1)}{(n - \frac{\sum_{s=1}^{r} n_s^2}{n})}$$

where $M_i$ is the variance within schools

$M_{bs}$ is the variance estimated from the means of schools

$n$ is the number of children

$n_s$ is the number of children in the $s^{th}$ school

$r$ is the number of schools

$$V_{bs} = \frac{1,254.8746 \times 45}{(2141 - \frac{101,833}{2141})}$$

$$= \frac{56469.357}{2093.4}$$

$$= \underline{26.975}$$

and the best estimate of the population variance is

$$M_i = V_i = 203.3476$$
$$V_{bs} + = \underline{26.975}$$
$$\underline{230.3226}$$

$\sigma = \underline{15.18}$ which is only slightly greater than the estimate obtained from the whole distribution, but somewhat greater than the 14.26 obtained from the within schools.

## The Standard Error of the Estimate of the Population Mean

From the estimates of the two components, i.e. within and between schools, it is a short step to estimate the standard error of the estimate of the population mean. It is assumed here that the 46 schools tested are a random selection from an infinite (or at least very large c. f. 46) number of schools associated with the City of Winnipeg, with a similar assumption with regard to the children tested in a given school. (It may be argued that the 2,141 cases are the total population in the particular age group, in which case the population variance is that of the 2,141 cases, and the population mean is known exactly, i.e. has no standard error.)

Again from the formulae derived in the appendix

$$V_{\bar{x}} = \frac{V_i}{n} + \frac{V_{bs} \times \sum_{s=1}^{r} n_s^2}{n^2} \qquad \text{Formula A}$$

when $V_{\bar{x}}$ is the variance of the estimate of the population mean.

$$= \frac{203.3476}{2141} + \frac{26.975 \times 47.6}{2141}$$

$$= \underline{.0950} + .6997$$

$$= \underline{.7947}$$

i.e. the standard error of the population mean obtained from the total distribution is .891. and since this estimate is 98.03 the weighted mean obtained by Stein, 99.38, does not differ significantly from it. However, 99.38 is the better estimate to use.

The technique used in this solution has been applied to data obtained by the Australian Council for Educational Research for the purpose of comparing the performance of students from the six states of Australia. The scores of 35,000 children from 600 schools have been examined and the standard errors of estimated State means (calculated as above) used to test the significance of the differences obtained. The possibility of using similar methods in factor studies based on heterogeneous populations is at present being examined.

The writer is indebted to E. J. Williams for suggesting the method of approach used in this article.

## References

1. H. L. Stein, "On Establishing Norms in a Large City School," Journal of Experimental Education, XVI (March 1948), pp. 192-195.

2. E. F. Lindquist, Statistical Analysis in Educational Research (Boston: Houghton Mifflin Co., 1940).

## Appendix

Derivation of Formulae:

Let $x_{is}$ represent the score of the ith person in the sth school.
$n_s$ the number of children in the sth school.
$r$ the number of schools.

Then the total number of children $n = \sum_{s=1}^{r} n_s$

the mean of the sth school $\bar{x}_s = \dfrac{\sum_{i=1}^{n_s} x_{is}}{n_s}$

the mean for the sample (estimate of the population mean)

$$\bar{\bar{x}} = \frac{\sum_{s=1}^{r} n_s \bar{x}_s}{n}$$

Let $V_i$ and $V_{bs}$ represent two components of the population variance -- $V_i$ due to differences between pupils within a school and $V_{bs}$ due to differences between schools caused by selective factors which vary from school to school. Then since these two factors are independent the population variance is equal to $V_i + V_{bs}$. Then since these two factors are independent the population variance is equal to $\underline{V_i + V_{bs}}$. Then $V_{\bar{x}_s} = \dfrac{V_i}{n_s} + V_{bs}$

$$V_{\bar{x}} = \frac{\sum_{s=1}^{r} n_s^2 V_{\bar{x}_s}}{n^2} = \frac{\sum_{s=1}^{r} n_s^2 \left(\frac{V_i}{n_s} + V_{bs}\right)}{n^2}$$

$$= \frac{V_i}{n} + \frac{V_{bs} \cdot \sum_{s=1}^{r} n_s}{n^2} \quad \text{(Formula A)}$$

To obtain formula for estimating $V_i$ and $V_{bs}$ we examine the expected values of the sums of squares within schools and between school means and equate these two expected values to the estimates of them obtained from the sample being examined.

Then $E(s_i)$ - the expected value of the pooled sum of squares within schools

$$= \sum_{s=1}^{r} (n_s - 1)\ V_i = \underline{(n - r)\ V_i}$$

i.e. $V_i = \dfrac{E(s_i)}{n - r}$ and hence an unbiased estimate of $V_i$ is the pooled within schools variance, the value used, erroneously I believe, by Stein for the population variance, and

$$E(S_{bs}) = \sum_{s=1}^{r} n_s V(\bar{x}_s) - n V_{\bar{x}}$$

$$= \sum_{s=1}^{r} (V_i + n_s V_{bs}) - V_i - V_{bs} \frac{\sum_{s=1}^{r} n_s^2}{n}$$

$$= V_i (r - 1) + V_{bs} \left(n - \frac{\sum n_s}{n}\right)$$

$$\therefore \quad V_{bs} = \frac{E(S\ ) - V\ (r - 1)}{\left(n - \dfrac{\sum n_s^2}{n}\right)}$$

Let $M_i$ represent the estimate of variance obtained by pooling the within schools variance.

$M_{bs}$ the estimate of variance obtained from the variance of the school means.

then $V_i = M_i$

$$V_{bs} = \frac{(M_{bs} - M_i)(r - 1)}{\left(n - \dfrac{\sum n_s^2}{n}\right)} \quad \text{(Formula B)}$$

If all the $n_s$ are equal (i.e. same number of children in each school) and equal $\bar{n}_s$ say, then $\bar{n}_s V_{bs} = M_{bs} - M_i$

and $V_{\bar{x}} = \dfrac{M_{bs}}{r.n_s} = \dfrac{M_{bs}}{n}$ (a formula recommended by E. F. Lindquist[2] but which should be used only when the condition of equal numbers in schools is satisfied.)

# A GRAPHICAL REPRESENTATION OF TEST RELIABILITY

JOHN A. JARVIS

Stout Institute

The Kuder-Richardson short formula* provides an easy means of graphically representing the relationship between the reliability of a test and the standard deviation of that test. This formula appears below:

$$r_{tt} = \frac{n}{n-1} \cdot \frac{\sigma^2 - npq}{\sigma^2}$$

where:

$r_{tt}$ = reliability of the test

$n$ = number of items

$\sigma$ = the standard deviation

$p$ = the mean divided by n

$q = 1 - p$

The variables p and q may be eliminated from the original formula in this manner:

let $x = npq$

but $q = 1 - p$

then $x = np(1 - p)$

$$x = np - np^2$$

Taking the first derivative of X with respect to p

$$\frac{dx}{dp} = n - 2np$$

Since npq is to be a maximum, let

$$\frac{dx}{dp} = 0$$

Then

$$0 = n - 2np$$

$$2np = n$$

$$p = \tfrac{1}{2} \text{ or } .5$$

$$q = 1 - .5$$

$$q = .5$$

The expression npq will then be a maximum when p and q are equal to .5. Substituting this value for p and q we have

$$r_{tt} = \frac{n}{n-1} \cdot \frac{\sigma^2 - n(.5)(.5)}{\sigma^2}$$

$$r_{tt} = \frac{n}{n-1} \cdot \frac{\sigma^2 - .25n}{\sigma^2}$$

Using the above formula it is possible to construct the family of curves included as Figure 1. This graph then provides a means of visualizing the increase in reliability with an increase in variability and a rough check on test reliability when computed by other means.

* G. F. Kuder and M. W. Richardson, "The Theory of the Estimation of Test Reliability," Psychometrika, II (September 1937), pp. 151-160.

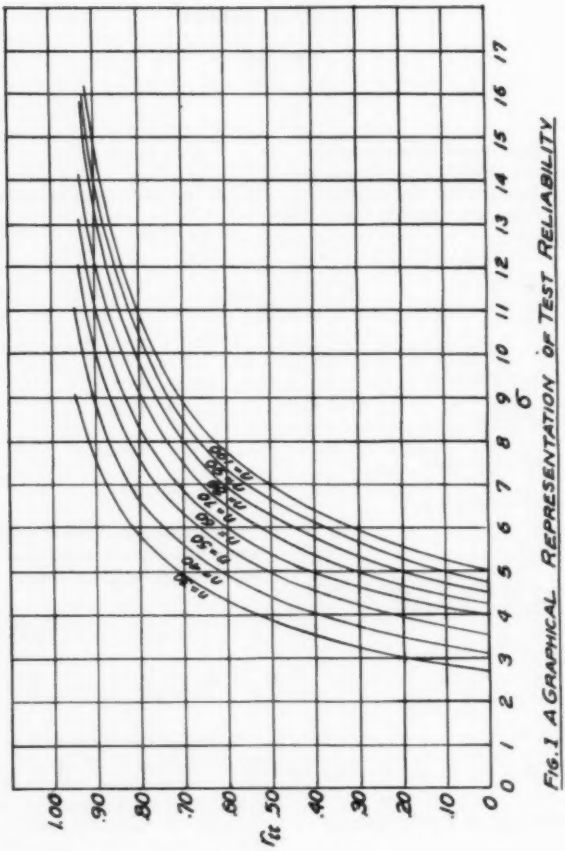


FIG. 1 A GRAPHICAL REPRESENTATION OF TEST RELIABILITY